



INVESTOR IN PEOPLE

EPO - DG 1

02 03. 2001

(78)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

Best Available Copy

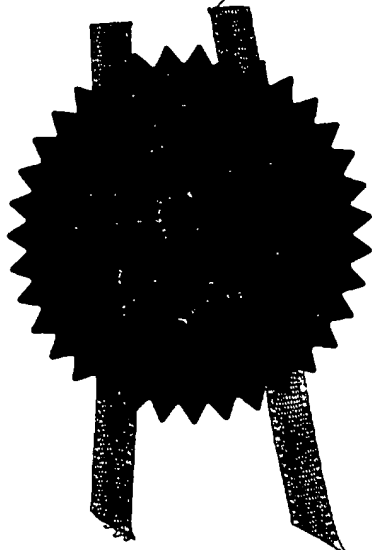
I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

I also certify that the attached copy of the request for grant of a Patent (Form 1/77) bears an amendment, effected by this office, following a request by the applicant and agreed to by the Comptroller-General.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Dated 22 February 2001

Patents Act 1977
(Rule 16)

The
Patent
Office

1/77

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)



17AUG98 E383546-2 D02862
P01/7700 25.00 - 9817796.7

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference SCB/50899/000

14 AUG 1998

2. Patent application number
(The Patent Office will fill in this part)

9817796.7

3. Full name, address and postcode of the or of each applicant (*underline all surnames*)

JANSSEN PHARMACEUTICA N.V.
Turnhoutseweg 30
B-2340 Beerse
BELGIUM

Patents ADP number (*if you know it*)

If the applicant is a corporate body, give the country/state of its incorporation

BELGIUM

005 31939001

4. Title of the invention DRUG TARGETS IN CANDIDA ALBICANS

5. Name of your agent (*if you have one*)

BOULT WADE TENNANT
27 FURNIVAL STREET
LONDON
EC4A 1PQ

"Address for service" in the United Kingdom to which all correspondence should be sent (*including the postcode*)

Verulam Gardens
70 Gray's Inn Road
London
WC1X 8BT

Patents ADP number (*if you know it*)

42001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (*if you know it*) the or each application number

Country

Priority application number
(if you know it)

Date of filing
(day/month/year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application

Date of filing
(day / month / year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request?

YES

(Answer 'Yes' if:

- a) any applicant named in part 3 is not an inventor, or
- b) there is an inventor who is not named as an applicant, or
- c) any named applicant is a corporate body.

See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description 27

Claim(s) 5

Abstract

Drawing(s) 74



10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

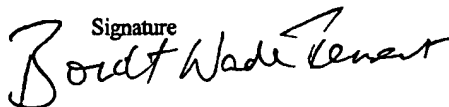
Statement of inventorship and right to grant of a patent (*Patents Form 7/77*)

Request for preliminary examination and search (*Patents Form 9/77*)

Request for substantive examination (*Patents Form 10/77*)

Any other documents
(Please specify)

11. I/We request the grant of a patent on the basis of this application.

Signature


Date

14 August 1998

12. Name and daytime telephone number of person to contact in the United Kingdom **COLM D. MURPHY**
0171 404 5921

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505.
- b) Write your answers in capital letters using black ink or you may type them.
- c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- e) Once you have filled in the form you must remember to sign and date it.
- f) For details of the fee and ways to pay please contact the Patent Office.

DRUG TARGETS IN CANDIDA ALBICANS

The present invention is concerned with the identification of genes or functional fragments thereof from *Candida albicans* which are critical for growth and cell division and which genes may be used as selective drug targets to treat *Candida albicans* associated infections. Novel nucleic acid sequences from *Candida albicans* are also provided and which encode the polypeptides which are critical for growth of *Candida albicans*.

Opportunistic infections in immunocompromised hosts represent an increasingly common cause of mortality and morbidity. *Candida* species are among the most commonly identified fungal pathogens associated with such opportunistic infections, with *Candida albicans* being the most common species. Such fungal infections are thus problematical in, for example, AIDS populations in addition to normal healthy women where *Candida albicans* yeasts represent the most common cause of vulvovaginitis.

Although compounds do exist for treating such disorders, such as for example, amphotericin, these drugs are generally limited in their treatment because of their toxicity and side effects. Therefore, there exists a need for new compounds which may be used to treat *Candida* associated infections in addition to compounds which are selective in their action against *Candida albicans*.

Classical approaches for identifying anti-fungal compounds have relied almost exclusively on inhibition of fungal or yeast growth as an endpoint. Libraries of natural products, semi-synthetic, or synthetic chemicals are screened for their ability to kill or arrest growth of the target pathogen or a related

nonpathogenic model organism. These tests are cumbersome and provide no information about a compounds mechanism of action. The promising lead compounds that emerge from such screens must then be tested for possible host-toxicity and detailed mechanism of action studies must subsequently be conducted to identify the affected molecular target.

The present inventors have now identified a range of nucleic acid sequences from *Candida albicans* which encode polypeptides which are critical for its survival and growth. These sequences represent novel targets which can be incorporated into an assay to selectively identify compounds capable of inhibiting expression of such polypeptides and their potential use in alleviating diseases or conditions associated with *Candida albicans* infection.

Therefore, according to a first aspect of the invention there is provided a nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides illustrated in Figures 1, 2, 4 to 7, 9 to 11, 13, 15 to 20, 22 to 26, 28 to 32, 34 to 43, 45a and b, 47 to 49, 51, 52, 53 to 57, 59 and 60.

A further aspect of the invention comprises a nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides illustrated in Figure 1, 2, 36, 37a and b, 38, 39 and 40 and fragments or derivatives of said nucleic acid molecules.

Letters utilised in the sequences according to the invention which are not recognisable as letters of the genetic code signify a position in the nucleic acid sequence where one or more of bases A, G, C or T

1

10

20

25

30

35

The nucleic acid capable of hybridising to nucleic acid molecules according to the invention will generally be at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the nucleotide sequences illustrated in any of Figures 1 to 61.

The DNA molecules according to the invention may, advantageously, be included in a suitable expression vector to express polypeptides encoded therefrom in a suitable host.

An expression vector according to the invention includes a vector having a nucleic acid according to the invention operably linked to regulatory sequences, such as promoter regions, that are capable of effecting expression of said DNA fragments. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. Such vectors may be transformed into a suitable host cell to provide for expression of a polypeptide according to the invention. Thus, in a further aspect, the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell, transformed or transfected with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides.

The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of said nucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable markers, such as, for example, ampicillin

resistance.

Polynucleotides according to the invention may be inserted into the vectors described in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA or other antisense nucleic acids may be produced by synthetic means.

In accordance with the present invention, a defined nucleic acid includes not only the identical nucleic acid but also any minor base variations including in particular, substitutions in bases which result in a synonymous codon (a different codon specifying the same amino acid residue) due to the degenerate code in conservative amino acid substitutions. The term "nucleic acid sequence" also includes the complementary sequence to any single stranded sequence given regarding base variations.

The present invention also comprises within its scope proteins or polypeptides expressed by the nucleic acid molecules according to the invention or a functional equivalent, derivative or bioprecursor thereof.

The present invention also advantageously provides nucleic acid sequences of at least approximately 15 contiguous nucleotides of a nucleic acid according to the invention and preferably from 15 to 50 nucleotides. These sequences may, advantageously be used as probes or primers to initiate replication, or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used in diagnostic kits or the like for detecting the presence of a nucleic acid according to the invention. These tests generally comprise contacting the probe with the sample under hybridising conditions and detecting for the presence

of any duplex or triplex formation between the probe and any nucleic acid in the sample.

Advantageously, the nucleic acid sequences, according to the invention may be produced using such recombinant or synthetic means, such as for example using PCR cloning mechanisms which generally involve making a pair of primers, which may be from approximately 15 to 50 nucleotides to a region of the gene which is desired to be cloned, bringing the primers into contact with mRNA, cDNA, or genomic DNA from a human cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolated the amplified region or fragment and recovering the amplified DNA. Generally, such techniques as defined herein are well known in the art, such as described in Sambrook et al (Molecular Cloning: a Laboratory Manual, 1989).

The nucleic acids or oligonucleotides according to the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S , enzyme labels or other protein labels such as biotin or fluorescent markers. Such labels may be added to the nucleic acids or oligonucleotides of the invention and may be detected using known techniques *per se*.

The polypeptide or protein according to the invention includes all possible amino acid variants encoded by the nucleic acid molecule according to the invention including a polypeptide encoded by said molecule and having conservative amino acid changes. Polypeptides according to the invention further include variants of such sequences, including naturally occurring allelic variants which are substantially homologous to said polypeptides. In

this context, substantial homology is regarded as a sequence which has at least 70%, preferably 80 or 90% amino acid homology with the polypeptides encoded by the nucleic acid molecules according to the invention.

5 Nucleic acids and polypeptides which are particularly preferred are those comprising the sequences of nucleotides illustrated in Figures 1 and 2. These sequences are specific to *Candida albicans* with no functionally related sequences in other
10 prokaryotic or eukaryotic organism as yet identified from the respective genomic databases.

 Nucleotide sequences according to the invention are particularly advantageous for selective
15 therapeutic targets for treating *Candida albicans* associated infections. For example, an antisense nucleic acid capable of binding to the nucleic acid sequence illustrated in any of Figures 1 to 61 may be used to selectively inhibit expression of the
20 corresponding polypeptides, leading to impaired growth of the *Candida albicans* with reductions of associated illnesses or diseases. The antisense nucleic acid corresponding to the sequences identified in Figures 1 and 2 may therefore be particularly useful in
25 selective treatment of *Candida albicans* associated infection.

 The nucleic acid molecule or the polypeptide according to the invention may be used as a
30 medicament, or in the preparation of a medicament, for treating diseases or conditions associated with *Candida albicans* infection.

 Advantageously, the nucleic acid molecule or the polypeptide according to the invention may be provided in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or
35 excipient therefor.

Antibodies to the protein or polypeptide of the present invention may, advantageously, be prepared by techniques which are known in the art. For example, polyclonal antibodies may be prepared by inoculating a
5 host animal, such as a mouse, with the polypeptide according to the invention or an epitope thereof and recovering immune serum. Monoclonal antibodies may be prepared according to known techniques such as described by Kohler R. and Milstein C., Nature (1975)
10 256, 495-497.

Antibodies according to the invention may also be used in a method of detecting for the presence of a polypeptide according to the invention, which method comprises reacting the antibody with a sample and
15 identifying any protein bound to said antibody. A kit may also be provided for performing said method which comprises an antibody according to the invention and means for reacting the antibody with said sample.

Proteins which interact with the polypeptide of the invention may be identified by investigating
20 protein-protein interactions using the two-hybrid vector system first proposed by Chien et al (1991).

This technique is based on functional reconstitution in vivo of a transcription factor which
25 activates a reporter gene. More particularly the technique comprises providing an appropriate host cell with a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating
30 domain, expressing in the host cell a first hybrid DNA sequence encoding a first fusion of a fragment or all of a nucleic acid sequence according to the invention and either said DNA binding domain or said activating domain of the transcription factor, expressing in the
35 host at least one second hybrid DNA sequence, such as

a library or the like, encoding putative binding proteins to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion;
5 detecting any binding of the proteins to be investigated with a protein according to the invention by detecting for the presence of any reporter gene product in the host cell; optionally isolating second hybrid DNA sequences encoding the binding protein.

10 An example of such a technique utilises the GAL4 protein in yeast. GAL4 is a transcriptional activator of galactose metabolism in yeast and has a separate domain for binding to activators upstream of the galactose metabolising genes as well as a protein
15 binding domain. Nucleotide vectors may be constructed, one of which comprises the nucleotide residues encoding the DNA binding domain of GAL4. These binding domain residues may be fused to a known protein encoding sequence, such as for example the
20 nucleic acids according to the invention. The other vector comprises the residues encoding the protein binding domain of GAL4. These residues are fused to residues encoding a test protein. Any interaction between polypeptides encoded by the nucleic acid
25 according to the invention and the protein to be tested leads to transcriptional activation of a reporter molecule in a GAL-4 transcription deficient yeast cell into which the vectors have been transformed. Preferably, a reporter molecule such as
30 β -galactosidase is activated upon restoration of transcription of the yeast galactose metabolism genes.

Further provided by the present invention is one or more *Candida albicans* cells comprising an induced mutation in the DNA sequence encoding the polypeptide
35 according to the invention.

A further aspect of the invention provides a method of identifying compounds which selectively inhibit expression of polypeptides expressed from the nucleotides sequences illustrated in any of Figures 1 to 61 and which are critical for growth and survival of *Candida albicans*, which method comprises (a) contacting a compound to be tested with one or more *Candida albicans* cells having a mutation in a nucleic acid molecule according to the invention which mutation results in overexpression or underexpression of said polypeptides in addition to one or more wild type *Candida* cells, (b) monitoring the growth and/or activity of said mutated cell compared to said wild type wherein differential growth or activity of said one or more mutated *Candida* cells provides an indication of selective action of said compound on said polypeptide or another polypeptide in the same or a parallel pathway.

Compounds identifiable or identified using the method according to the invention, may advantageously be used as a medicament, or in the preparation of a medicament to treat diseases or conditions associated with *Candida albicans* infection. These compounds may also advantageously be included in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

A further aspect of the invention provides a method of identifying DNA sequences from a cell or organism which DNA encodes polypeptides which are critical for growth or survival, which method comprises (a) preparing a cDNA or genomic library from said cell or organism in a suitable expression vector which vector is such that it can either integrate into the genome in said cell or that it permits transcription of antisense RNA from the nucleotide

sequences in said cDNA or genomic library, (b)
selecting transformants exhibiting impaired growth and
determining the nucleotide sequence of the cDNA or
genomic sequence from the library included in the
5 vector from said transformant. Preferably, the cell
or organism may be any yeast or filamentous fungi,
such as, for example, *Saccharomyces cerevisiae*,
Saccharomyces pombe or *Candida albicans*.

A further aspect of the invention provides a
10 pharmaceutical composition comprising a compound
according to the invention together with a
pharmaceutically acceptable carrier, diluent or
excipient therefor.

The present invention may be more clearly
15 understood with reference to the accompanying example,
which is purely exemplary, with reference to the
accompanying drawings, wherein

20 Figures 1 & 2: are nucleotide sequences of
previously unknown function
isolated from *Candida albicans* and
which sequences are not present in
the public domain.

25 Figures 3 to 35: are nucleotide sequences of
previously unknown function
isolated from *Candida albicans* and
which sequences are partially or
30 fully present in the public
domain.

35 Figures 36 to 40: are nucleotide sequences isolated
from *Candida albicans* and which
have an identified function based
on sequence homology with proteins

from other organisms and which sequences are not present in the public domain.

5 Figures 41 to 61: are nucleotide sequences having an identified function based on sequence homology comparisons from other organisms and which sequences are fully or partially present in the public domain.

10 Figure 62: is a diagrammatic representation of plasmid pGAL1PNiST-1.

15 Figure 63: is a nucleotide sequence of plasmid pGAL1PNiST-1 of Figure 62.

 Figure 64: is a diagrammatic representation of plasmid pGAL1PSiST-1.

20 Figure 65: is a nucleotide sequence of plasmid pGAL1PSiST-1 of Figure 64.

25 Figures 66 to 106: are amino acid sequences of the appropriately corresponding DNA sequences illustrated in Figures 1 to 61.

30 Example 1

 Identification of novel drug targets in *C. albicans* by anti-sense and disruptive integration

 The principle of the approach is based on the fact that when a particular *C. albicans* mRNA is inhibited by producing the complementary anti-sense

35

RNA, the corresponding protein will decrease. If this protein is critical for growth or survival, the cell producing the anti-sense RNA will grow more slowly or will die.

5 Since anti-sense inhibition occurs at mRNA level, the gene copy number is irrelevant, thus allowing applications of the strategy even in diploid organisms.

10 Anti-sense RNA is endogenously produced from an integrative or episomal plasmid with an inducible promoter; induction of the promoter leads to the production of a RNA encoded by the insert of the plasmid. This insert will differ from one plasmid to another in the library. The inserts will be derived
15 from genomic DNA fragments or from cDNA to cover to the extent possible- the entire genome.

 The vector is a proprietary vector allowing integration by homologous recombination at either the homologous insert or promoter sequence in the *Candida*
20 genome. After introducing plasmids from cDNA or genomic libraries into *C. albicans*, transformants are screened for impaired growth after promoter (& thus anti-sense) induction in the presence of lithium acetate. Lithium acetate prolongs the G1 phase and
25 thus allows anti-sense to act during a prolonged period of time during the cell cycle. Transformants which show impaired growth in both induced and non-induced media, thus showing a growth defect due to integrative disruption, are selected as well.

30 Transformants showing impaired growth are supposed to contain plasmids which produce anti-sense RNA to mRNAs critical for growth or survival. Growth is monitored by measuring growth-curves over a period of time in a device (Bioscreen Analyzer, Labsystems)
35 which allows simultaneous measurement of growth-curves

of 200 transformants.

Subsequently plasmids can be recovered from the transformants and the sequence of their inserts determined, thus revealing which mRNA they inhibit. In order to be able to recover the genomic or cDNA insert which has integrated into the *Candida* genome, genomic DNA is isolated, cut with an enzyme which cuts only once into the library vector (and estimated approx. every 4096 bp in the genome) and religated. PCR with primers flanking the insert will yield (partial) genomic or cDNA inserts as PCR fragments which can directly be sequenced. This PCR analysis (on ligation reaction) will also show us how many integrations occurred. Alternatively the ligation reaction is transformed to *E. coli* and PCR analysis is performed on colonies or on plasmid DNA derived thereof.

This method is employed for a genome wide search for novel *C. albicans* genes which are important for growth or survival.

Materials & Methods

Construction of pGAL1PNiST-1

The backbone of the pGAL1PNiST-1 vector (integrative anti-sense *SfiI*-*NotI* vector) is pGEM11Zf(+) (Promega Inc.). First, the CaMAL2 *EcoRI*/*SalI* promoter fragment from pDBV50 (D.H. Brown et al.) was ligated into *EcoRI*/*SalI*-opened pGEM11Zf(+) resulting in the intermediate construct pGEMMAL2P-1. Into the latter (*MscI*/*CIP*) the CaURA3 selection marker was cloned as a *Eco47III*/*XmnI* fragment derived from pRM2. The resulting pGEMMAL2P-2 vector was *NotI*/*HindIII* opened in order to accept the *NotI*-stuffer-*SfiI* cassette from pPCK1NiSCYCT-1 (*EagI*/*HindIII* fragment): pMAL2PNiST-1. Finally, the plasmid pGAL1PNiST-1 was constructed by exchanging the

SalI/*Ecl136II* MAL2 promoter in pMAL2PNiST-1 by the *XhoI*/*SmaI* GAL1 promoter fragment derived from pRM2GAL1P.

5 **Construction of pGAL1PSiST-1**

 The vector pGAL1PSiST-1 was created for cloning the small genomic DNA fragments (flanked by *SfiI* sites) behind the GAL1 promoter. The only difference with pGAL1PNiST-1 is that the hIFN β (stuffer fragment) insert fragment in pGAL1PSiST-1 is flanked by two *SfiI* sites in stead of a *SfiI* and a *NotI* site as in pGAL1PNiST-1. To construct pGAL1PSiST-1 the *EcoRI*-*HindIII* fragment, containing hIFN β flanked by a *SfiI* and a *NotI* site, of pMAL2pHiET-3 (unpublished) was
10 exchanged by the *EcoRI*-*HindIII* fragment, containing hIFN β flanked by two *SfiI* sites, from YCp50S-S (an *E. coli* / *S. cerevisiae* shuttle vector derived from the plasmid YCp50, which is deposited in the ATCC collection (number 37419; Thrash et al., 1985); an
15 *EcoRI*-*HindIII* fragment, containing the gene hIFN β , which is flanked by two *SfiI* sites, was inserted in YCp50, creating YCp50S-S), resulting into plasmid pMAL2PSiST-1. The *mal2* promoter from pMAL2PSiST-1 (by a *NaeI*-*FspI* digest) was further replaced by the *gal1*
20 promoter from pGAL1PNiST-1 (via a *XhoI*-*SalI* digest), creating the vector pGAL1PSiST-1.
25

***Candida albicans* genomic library**

 * *Preparation of the genomic DNA fragments*

30 A *Candida albicans* genomic DNA library with small DNA fragments (400 to 1,000 bp) was prepared. Genomic DNA of *Candida albicans* B2630 was isolated following a modified protocol of Blin and Stafford (1976). The quality of the isolated genomic DNA was checked by gel
35 electrophoresis. Undigested DNA was located on the gel

above the marker band of 26,282 bp. A little smear, caused by fragmentation of the DNA, was present. To obtain enrichment for genomic DNA fragments of the desired size, the genomic DNA was partially digested.

5 Several restriction enzymes (*AluI*, *HaeIII* and *RsaI*; all creating blunt ends) were tried out. The appropriate digest conditions have been determined by titration of the enzyme. Enrichment of small DNA fragments was obtained with 70 units of *AluI* on 10 µg

10 of genomic DNA for 20 min. T4 DNA polymerase (Boehringer) and dNTPs (Boehringer) were added to polish the DNA ends. After extraction with phenol-chloroform the digest was size-fractionated on an agarose gel. The genomic DNA fragments with a length

15 of 500 to 1,250 bp were eluted from the gel by centrifugal filtration (Zhu et al., 1985). *SfiI* adaptors (5' GTTGGCCTTTT) or (5' AGGCCAAC) were attached to the DNA ends (blunt) to facilitate cloning of the fragments into the vector. Therefore, a 8-mer

20 and 11-mer oligonucleotide (comprising the *SfiI* site) were kinased and annealed. After ligation of these adaptors to the DNA fragments a second size-fractionation was performed on an agarose gel. The DNA fragments of 400 to 1150 bp were eluted from the

25 gel by centrifugal filtration.

* Preparation of the *pGAL1PSiST-1* vector fragment

The small genomic DNA fragments were cloned after the *GAL1* promoter in the vector *pGAL1PSiST-1*. Qiagen-purified *pGAL1PSiST-1* plasmid DNA was digested with

30 *SfiI* and the largest vector fragment eluted from the gel by centrifugal filtration (Zhu et al., 1985). Ligation with a control DNA fragment, flanked by *SfiI* sites, was performed as a control. The ligation mix was electroporated to MC1061 *E. coli* cells. Plasmid

35 DNA of 24 clones was analyzed. In all cases the

control fragment was inserted in the pGAL1PSiST-1 vector fragment.

* *Upscaling*

5 All genomic DNA fragments (450 ng) were ligated into the pGAL1PSiST-1 vector (20 ng). After electroporation at 2500V, 40 μ F circa 400,000 clones were obtained. These clones were pooled into three groups and stored as glycerol slants. Also Qiagen-
10 purified DNA was prepared from these clones. A clone analysis showed an average insert length of 600 bp and a percentage of 91 for clones with an insert. The size of the library corresponds to 5 times the diploid genome. The genomic DNA inserts are sense or anti-sense orientated in the vector.
15

Candida albicans cDNA library

Total RNA was extracted from *Candida albicans* B2630 grown on respectively minimal (SD) and rich
20 (YPD) medium as described by Chirgwin et al in Sambrook et al. mRNA was prepared from total RNA using the Invitrogen Fast Track procedure.

First strand cDNA is synthesised with the Superscript Reverse Transcriptase (BRL) and with an
25 oligo dT-NotI Primer adapter. After second strand synthesis, cDNA is polished with Klenow enzyme and purified over a Sephacryl S-400 spun column. Phosphorylated *Sfi*I adapters are then ligated to the cDNA, followed by digestion with the NotI restriction
30 enzyme. The *Sfi*I/NotI cDNA is then purified and sized on a Biogel column A150M.

First fraction contains approximately 38,720 clones by transformation, the second fraction only 1540 clones. Clone analysis:
35 Fr. I: 22/24 inserts, 16 \geq 1000 bp, 4 \geq 2000 bp,

average size: 1500 bp.

Fr. II: 9/12 inserts, 3 \geq 1000 bp, average size: 960 bp cDNA was ligated in a *NotI/SfiI* opened pGAL1PNiST-1 vector (anti-sense)

5

Candida transformation

The host strain used for transformation is a *C. albicans* *ura3* mutant, CAI-4, which contains a deletion in orotidine-5'-phosphate decarboxylase and was
10 obtained from William Fonzi, Georgetown University (Fonzi and Irwin). CAI-4 was transformed with the above described cDNA library or genomic library using the *Pichia* spheroplast module (Invitrogen). Resulting transformants were plated on minimal medium
15 supplemented with glucose (SD, 0.67% or 1.34% Yeast Nitrogen base w/o amino acids + 2% glucose) plates and incubated for 2-3 days at 30°C.

Screening for mutants

20 Starter cultures were set up by inoculating each colony in 1 ml SD medium and incubating overnight at 30°C and 300 rpm. Cell densities were determined using a Coulter counter (Coulter Z1; Coulter electronics limited). 250,000 cells/ml were inoculated in 1 ml SD
25 medium and cultures were incubated for 24 hours at 30°C and 300 rpm. Cultures were washed in minimal medium without glucose (S) and the pellet resuspended in 650 μ l S medium. 8 μ l of this culture is used for inoculating 400 μ l cultures in a Honeywell-100 plate
30 (Bioscreen analyzer; Labsystems). Each transformant was grown during three days in S medium containing LiAc; pH 6.0, with 2% glucose/2% maltose or 2% galactose/2% maltose respectively while shaking every 3 minutes for 20 seconds. Optical densities were
35 measured every hour during three consecutive days and

growth curves were generated (Bioscreen analyzer; Labsystems).

Growth curves of transformants grown in respectively anti-sense non-inducing (glucose/maltose) and inducing (galactose/maltose) medium are compared and those transformants showing impaired growth upon anti-sense induction are selected for further analysis. Transformants showing impaired growth by virtue of integration into a critical gene are also selected.

Isolation of genomic or cDNA inserts

Putatively interesting transformants are grown in 1.5 ml SD overnight and genomic DNA is isolated using the Nucleon MI Yeast kit (Clontech). Concentration of genomic DNA is estimated by analyzing a sample on an agarose gel.

20 ng of genomic DNA is digested for three hours with an enzyme that cuts uniquely in the library vector (SacI for the genomic library; PstI for the cDNA library) and treated with RNase. Samples are phenol/chloroform extracted and precipitated using NaOAc/ethanol.

The resulting pellet is resuspended in 500 μ l ligation mixture (1 x ligation buffer and 4 units of T4 DNA ligase; both from Boehringer) and incubated overnight at 16°C.

After denaturation (20 min 65°C), purification (phenol/chloroform extraction) and precipitation (NaOAc/ethanol) the pellet is resuspended in 10 μ l MilliQ (Millipore) water.

PCR analysis

Inverse PCR is performed on 1 μ l of the precipitated ligation reaction using library vector

specific primers (oligo23 5' TGC-AGC-TCG-ACC-TCG-ACT-G
3' and oligo25 5' GCG-TGA-ATG-TAA-GCG-TGA-C 3' for the
genomic library; 3pGALNistPCR primer

:5'TGAGCAGCTCGCCGTCGCGC 3' and 5pGALNistPCR primer:

5 5'GAGTTATACCCTGCAGCTCGAC 3' for the cDNA library; both
from Eurogentec) for 30 cycles each consisting of (a)
1 min at 95 °C, (b) 1 min at 57 °C, and (c) 3 min at
72 °C. In the reaction mixture 2.5 units of Taq
polymerase (Boehringer) with TaqStart antibody
10 (Clontech) (1:1) were used, and the final
concentrations were 0.2 µM of each primer, 3 mM MgCl₂
(Perkin Elmer Cetus) and 200 µM dNTPs (Perkin Elmer
Cetus). PCR was performed in a Robocycler
(Stratagene).

15

Sequence determination

Resulting PCR products were purified using PCR
purification kit (Qiagen) and were quantified by
comparison of band intensity on EtBr stained agarose
20 gel with the intensity of DNA marker bands. The amount
of PCR product (expressed in ng) used in the
sequencing reaction is calculated as the length of the
PCR product in basepairs divided by 10. Sequencing
reactions were performed using the ABI Prism BigDye
25 Terminator Cycle Sequencing Ready Reaction Kit
according to the instructions of the manufacturer (PE
Applied Biosystems, Foster City, CA) except for the
following modifications.

The total reaction volume was reduced to 15 µl.
30 Reaction volume of individual reagents were changed
accordingly. 6.0 µl Terminator Ready Reaction Mix was
replaced by a mixture of 3.0 µl Terminator Ready
Reaction Mix + 3.0 µl Half Term (GENPAK Limited,
Brighton, UK). After cycle sequencing, reaction
35 mixtures were purified over Sephadex G50 columns

prepared on Multiscreen HV opaque microtiter plates (Millipore, Molsheim, Fr) and were dried in a speedVac. Reaction products were resuspended in 3 μ l loading buffer. Following denaturation for 2 min at 95°C, 1 μ l of sample was applied on a 5% Long Ranger Gel (36 cm well-to-read) prepared from Singel Packs according to the supplier's instructions (FMC BioProducts, Rockland, ME). Samples were run for 7 hours 2X run on a ABI 377XL DNA sequencer. Data collection version 2.0 and Sequence analysis version 3.0 (for basecalling) software packages are from PE Applied Biosystems. Resulting sequence text files were copied onto a server for further analysis.

Sequence analysis

Nucleotide sequences were imported in the VectorNTI software package (InforMax Inc, North Bethesda, MD, USA), and the vector and insert regions of the sequences were identified. Sequence similarity searches against public and commercial sequence databases were performed with the BLAST software package (Altschul et al., 1990) version 1.4. Both the original nucleotide sequence and the six-frame conceptual translations of the insert region were used as query sequences. The used public databases were the EMBL nucleotide sequence database (Stoesser et al., 1998), the SWISS-PROT protein sequence database and its supplement TrEMBL (Bairoch and Apweiler, 1998), and the ALCES *Candida albicans* sequence database (Stanford University, University of Minnesota). The commercial sequence databases used were the LifeSeq® human and PathoSeq™ microbial genomic databases (Incyte Pharmaceuticals Inc., Palo Alto, CA, USA), and the GENESEQ patent sequence database (Derwent, London, UK). Three major results were obtained on the basis of

the sequence similarity searches: function, novelty, and specificity. A putative function was deduced on the basis of the similarity with sequences with a known function, the novelty was based on the absence
5 or presence of the sequences in public databases, and the specificity was based on the similarity with vertebrate homologues.

Methods

10 Blastx of the nucleic acid sequences against the appropriate protein databases: Swiss-Prot for clones of which the complete sequence is present in the public domain, and paorfp (PathoSeq™) for clones of which the complete sequences is not present in the
15 public domain.

The protein to which the translated nucleic acid sequence corresponds to is used as a starting point. The differences between this protein and our translated nucleic acid sequences are marked with a
20 double line and annotated above the protein sequence. The following symbols are used:

a one-letter amino acid code or the ambiguity code X is used if our translated nucleic acid sequence has another amino acid on a certain position,

25 the stop codon sign * is used if our translated nucleic acid sequence has a stop codon on a certain position,

The letters fs (frame shift) are used if a frame shift occurs in our translated nucleic acid sequence, and another reading frame is used,
30

the words ambiguity or ambiguities are used if a part of our translated nucleic acid sequence is present in the proteins, but not visible in the alignments of the blast results,

35 The phrase missing sequence is used if the

translated nucleic acid sequence does not comprise that part of the protein.

Blastx: compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database.

Screening for compounds modulating expression of polypeptides critical for growth and survival of *C. albicans*

The method proposed is based on observations (Sandbaken et al., 1990; Hinnebusch and Liebman 1991; Ribogene PCT WO 95/11969, 1995) suggesting that underexpression or overexpression of any component of a process (e.g. translation) could lead to altered sensitivity to an inhibitor of a relevant step in that process. Such an inhibitor should be more potent against a cell limited by a deficiency in the macromolecule catalyzing that step and/or less potent against a cell containing an excess of that macromolecule, as compared to the wild type (WT) cell.

Mutant yeast strains, for example, have shown that some steps of translation are sensitive to the stoichiometry of macromolecules involved. (Sandbaken et al.). Such strains are more sensitive to compounds which specifically perturb translation (by acting on a component that participates in translation) but are equally sensitive to compounds with other mechanisms of action.

This method thus not only provides a means to identify whether a test compound perturbs a certain process but also an indication of the site at which it exerts its effect. The component which is present in altered form or amount in a cell whose growth is affected by a test compound is potentially the site of action of the test compound.

The assay to be set up involves measurement of growth of an isogenic strain which has been modified only in a certain specific allele, relative to a wild type (WT) *C. albicans* strain, in the presence of R-
5 compounds. Strains can be ones in which the expression of a specific essential protein is impaired upon induction of anti-sense or strains which carry disruptions in an essential gene. An in silico approach to finding novel essential genes in *C.*
10 *albicans* will be performed. A number of essential genes identified in this way will be disrupted (in one allele) and the resulting strains can be used for comparative growth screening.

15 **Assay for High Throughput screening for drugs**
35 μ l minimal medium (S medium + 2% galactose + 2% maltose) is transferred in a transparent flat-bottomed 96 well plate using an automated pipetting system (Multidrop, Labsystems). A 96-channel pipettor
20 (Hydra, Robbins Scientific) transfers 2.5 μ l of R-compound at 10^{-3} M in DMSO from a stock plate into the assay plate.

The selected *C. albicans* strains (mutant and parent (CAI-4) strain) are stored as glycerol stocks
25 (15%) at -70°C . The strains are streaked out on selective plates (SD medium) and incubated for two days at 30°C . For the parent strain, CAI-4, the medium is always supplemented with 20 $\mu\text{g/ml}$ uridine. A single colony is scooped up and resuspended in 1 ml minimal
30 medium (S medium + 2% galactose + 2% maltose). Cells are incubated at 30°C for 8 hours while shaking at 250 rpm. A 10 ml culture is inoculated at 250,000 cells/ml. Cultures are incubated at 30°C for 24 hours while shaking at 250 rpm. Cells are counted in
35 Coulter counter and the final culture (S medium + 2%

galactose + 2% maltose) is inoculated at 20.000 to 50.000 cells/ml. Cultures are grown at 30°C while shaking at 250 rpm until a final OD of 0.24 (+/- 0.04) 6nM is reached.

- 5 200 µl of this yeast suspension is added to all wells of MW96 plates containing R-compounds in a 450 µl total volume. MW96 plates are incubated (static) at 30°C for 48 hours.

Optical densities are measured after 48 hours.

- 10 Test growth is expressed as a percentage of positive control growth for both mutant (x) and wild type (y) strains. The ratio (x/y) of these derived variables is calculated.

15

References

- Thrash C., Bankier A. T., Barrell B. G. and Sternglanz
R. (1985) *Proc. Natl. Acad. Sci. USA* 82: 4374-4378.
- 5 Blin N. and Stafford D. W. (1976) *Nucleic Acids Res.*
3: 2303-2308.
- 10 Zhu J., Kempnaers W., Van Der Straeten D., Contreras
R., and Fiers W. (1985) *Bio/Technology* 3: 1014-1016.
- 15 Sambrook J., Fritsch E.F. and Maniatis T. (1989)
Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold
Spring Harbor Laboratory, Cold Spring Harbor, NY.
- 20 Fonzi W. and Irwin H. (1993) *Genetics* 134:717-728.
Altschul S.F, Gish W., Miller W., Myers E.W., Lipman
D.J. (1990) *J. Mol. Biol.* 215(3):403-410.
- 25 Bairoch A. and Apweiler R. (1998) *Nucleic Acids Res.*
26(1):38-42.
- 30 Stoesser G., Moseley M.A., Sleep J., McGowran M.,
Garcia-Pastor M., Sterk P. (1998) *Nucleic Acids Res.*
26(1):8-15.
- 35 Chien et al., (1991) *Proc. Natl. Acad. Sci USA* 88,
9578-9582.
- 30 Sandbaken M.G. Lupisella J.A., DiDomenico B. and
Chakraborty K., *J. Biol. Chem.* 265:15838-15844, 1990.
- 35 Hinnebusch A.G. and Liebman S.W., in : *The Molecular
Biology of the Yeast Saccharomyces*, Broach J.R.,
Pringle J.R. and Jones E.W., eds., CSH Laboratory

Press; NY 1991.

Patent application RiboGene Inc., PCT WO 95./11969,
1995.

5

Brown, D.H. et al., Molecular General Genetics 251 (1)
75-80, 1996.

Claims

1. A nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides illustrated in Figure 1, 2, 4 to 7, 9 to 11, 13, 15 to 20, 22 to 26, 28 to 32, 34 to 43, 45a and b, 47 to 49, 51, 52, 53 to 57, 59 and 60.
2. A nucleic acid molecule encoding a polypeptide which is critical for survival and growth of the yeast *Candida albicans* and which nucleic acid molecule comprises any of the sequences of nucleotides illustrated in Figure 1, 2, 36, 37a, 38, 39 and 40 and fragments or derivatives of said nucleic acid molecules
3. A nucleic acid molecule according to claim 1 or 2 which is mRNA.
4. A nucleic acid molecule according to claim 1 or 2 which is DNA.
5. A nucleic acid molecule according to claim 4 which is cDNA.
6. A nucleic acid molecule capable of hybridising to the molecules according to any of claims 1 to 5 or the sequences illustrated in any of Figures 1 to 61 under high stringency conditions.
7. A polypeptide encoded by the nucleic acid molecule according to any of claims 1 to 6 or the sequence illustrated in any of Figures 1 to 61.

8. An expression vector comprising a nucleic acid molecule according to claim 4 or 5.

5 9. An expression vector according to claim 8 which comprises an inducible promoter.

10 10. An expression vector according to claim 8 or 9 which comprises a sequence encoding a reporter molecule.

11. A nucleic acid molecule according to any of claims 1 to 6 or the nucleotide sequences illustrated in Figure 1 to 61 for use as a medicament.

15 12. Use of a nucleic acid molecule according to any of claims 1 to 5 or the sequences illustrated in Figure 1 to 61 in the preparation of a medicament for treating *Candida albicans* associated diseases.

20 13. A polypeptide according to claim 7 for use as a medicament.

25 14. Use of a polypeptide according to claim 7 in the preparation of a medicament for treating *Candida albicans* associated infections.

30 15. A pharmaceutical composition comprising a nucleic acid molecule according to any of claims 1 to 6 or a polypeptide according to claim 7 together with a pharmaceutically acceptable carrier diluent or excipient therefor.

35 16. A *Candida albicans* cell comprising an induced mutation in the DNA sequence encoding the polypeptide according to claim 7.

17. A method of identifying compounds which selectively modulate expression of polypeptides which are crucial for growth and survival of *Candida albicans*, which method comprises:

- 5 (a) contacting a compound to be tested with one or more *Candida albicans* cells having a mutation in a nucleic acid molecule according to any of claims 1 to 5 which mutation results in overexpression or
10 underexpression of said polypeptides in addition to contacting one or more wild type *Candida albicans* cells with said compound,
 (b) monitoring the growth and/or activity of said mutated cell compared to said wild
15 type; wherein differential growth or activity of said one or more mutated *Candida* cells is indicative of selective action of said compound on a polypeptide or another polypeptide in the same or a parallel
20 pathway.

18. A compound identifiable according to the method of claim 17.

25 19. A compound according to claim 18 for use as a medicament.

30 20. Use of a compound according to claim 18 in the preparation of a medicament for treating *Candida albicans* associated diseases.

35 21. A pharmaceutical composition comprising a compound according to claim 18 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

22. A method of identifying DNA sequences from a cell or organism which DNA encodes polypeptides which are critical for growth or survival of said cell or organism, which method comprises:

- 5 (a) preparing a cDNA or genomic library from said cell or organism in a suitable expression vector which vector is such that it can either integrate into the genome in said cell or that it permits transcription of antisense RNA from the nucleotide sequences in said cDNA or genomic library.
- 10 (b) selecting transformants exhibiting impaired growth and determining the nucleotide sequence of the cDNA or genomic sequence from the library included in the vector from said transformant.
- 15

23. A method according to claim 22 wherein said cell or organism is a yeast or filamentous fungi.

20

24. A method according to claim 22 or 23 wherein said cell or organism is any of *Saccharomyces cervisiae*, *Saccharomyces pombe* or *Candida albicans*.

25 25. Plasmid pGAL1PSiST-1 having the sequence of nucleotides illustrated in Figure 63.

26. Plasmid pGAL1PNiST-1 having the sequence of nucleotides illustrated in Figure 65.

30

27. An antibody capable of binding to a polypeptide according to claim 7.

28. An oligonucleotide comprising a fragment of from 15 to 50 contiguous nucleic acid sequences of a

35

- 32 -

nucleic acid molecule according to any of claims 1 to
6.

Sequences with unknown function, *C. albicans* sequence NOT present in the public domain (ALCES/EMBL)

>328c2 1803bp in-house: 1123-1803 public: 1-436/468-1021 PathoSeq:
437-467/1022-1122

ATGTCTATTACAGTTACATTTCCGAAATCTCCATCTACGAAAAACGTGCACCG
GCATTTGGAATTGAGTTGGAGTTYAG
TCAMCAAGSCAGTAGCGATGGTGCTATAGAGAAAAGCGGCATTGGCAGTTCCT
GTGTTTAGCGTTGACAACCAAGACTWT
GTATTKATAAGAGAYCWTGCCAAGTACTGGGGCTACCCTTCATCGTATCAATT
GATTGTCAAGTTGGTCAAATGTGCTAA
CATTGAAAAGTCGCAAATCTTAAAGACCGATAAGGATTTGAATAGAGAGTTGT
TTGAGTTGGATTTGATTGAAGAAGCAG
ATACAAAGATTGATCTTTTTTATATTTTCGTTACCCTTGGTCTATTCAAGAATAGA
AAATAAGAAGGTTTTTTATGTTCTG
CGTGAACCAGAACAGCCAAAGGTGTGCGAAAGCMCCAACACAAGAGAAACCAG
CAAGTGTGGTTGCTGCAGAAGAAGATGA
CGATAATCTAGATGATGATGAGGAGGACGAAGTGGATGAAGACATGGATGAA
GATAATGATAATAGTGGGGAATTGTCTA
AAGGATACAAGCACATGCACAAGGACCATCCAAAGTATATAAATGACGATAG
GGTTACTATTGGACAAGTGTTTCATCAA
TACGGACTTGACCCTTCGACACCATTAAACCCATTCACTTTTCAATAGTATCAAC
TCAATGTGCGAAGCTAAACTATTACAA
GAATTTTGGAGTTTCAGGTTACCGATTTCTTCCCAACAGCAAGTTATCTTATGC
AGAACGAGAATTGGTGTGGAATGCCA
ACAACCTACAATGATATGCACATTAACGAAAAGACAGAATCCAAGCCGAAAAA
GAGTTTCCGTAAACCCATTGGAAAGTCA
AAGAAACATAACTTGCAGATTGATCCGAACTCCATAGATTTAAGCGAGTCAGT
GATTCCGGGACAAGGGTTTATACCTGA
CTTTAGTATCCACCTATCTTTGCAAAGTCCCTAATTATTATGTGACATCAACCC
ACCAAAGTCTCCCGCTGTCGTTCAAC
ACAAAGAATCTTAATGCAACTTCGAACTCTTCGTATTTGTTTAATGATAATGTC
AAGATAAAGTCAAAAAGTATTCAAGAA
GTWSGTGTTCAACAGCGATACCGATAATTACCATCACACAAAGTATTTCTACA
CCAAAACCTACCGTGGTCCAGGGTCGG
GGAATTACAAGGATGGTGCATTGATGAACAAAATCAACAAGATACATCTTTCC
AGTAATAAAAAGCCGCGCCACAAGAGA
AAGGTGTGCAACAATAACAGGTACAACAAGAGTTTAAAGGGGTTAGTCCACG
AAAAGTTTGACAAGAAGCTTTGTTGAGTA
CTTGCTTTCTGAGCAACGCAAGTATACCGAGGACTATTCCAATCTTGAAATTTT
ACACAATAGCTTACAGTTTAATGTTT
TTTTGAATACGTATCGTGGTGTGCCCCAAGAGACATGGAATAACTACTACAAG
TTTAAATTGATTGATTTTCAACAATTG
AAGGCTTTGCAAATGGAGGCAAATGAGCTTGAGGAGAGAAAATTGGATGCTG
CTAGACACCAACAGTGGGCGGAAGAAGA
GAAGCTTTNCCAAGAAAGATTGCGTTTAGTATTTGAAGATGAACGGACGAGTT
TGAGCAATTGCAAAGCGAGTTTGGTCA

Fig 1

GAGAAAGAAGGATTTGGAAGAGAAATTGCGTCGCCGTCAGCTANANGCATCTT
TGANTGATAGTTTTGAACTTGATAGCG
AAAATGACNATGAATCTTGACTTGNCCAAANTNAACAAGACTT

Fig 1 (cont)

>214c_cpL1 290bp inhouse:1-290

GAATCNCANACTCGNCACNGCTCCCCAAAAAGGCCAACGTTTCGTGCAAAAGGC
TATACTGGTGATATCCACGCAGATGAA
GAGCAAGTTTAATCAACTCTTTGTCAATTAATGCTGTACTTGTTCATTTTATT
TGCTGGCATTAAAGAATACCCATA
GTTTCAGAAAATAAAATTGAAAAATTTAAAAAAAACGCAATATCATTTCATTTT
TTTTGTTTTTTTGACAATAATATTAAT
ATGTAGTTACCAATGTTTTTAGATTTTATATGTTTTGAAAAAATAGTTTG

Fig 2

Sequences with unknown function, *C. albicans* sequence present in the public domain (ALCES)

>113g2 638bp in-house: 1-638

CTTATTCCGTTCTAGTGTCTCAATTGGTTATCCATTAACATCTATTCCCAACTCC
ATCATTATTGGCAATAAATAAATGG
GTGTTATATCTATTGGTAATAACTAACTGGTGTCAATTCAATTCCAATATGGT
CATGACAATTGAAAGTGTTACTGTTT
TGGTTTACATATTCTACAGGTTACAACCTATTGATTGGTTAGAAGTTTGGTTTCA
ACATCACCTGTTGCTAAGAATAAATG
TTGGTCATATCAATTGAATCATTTGTTGGTGTATGGTAAGTAAATGCTGGTTA
TATCTATTATCTACAACCACCAAGT
ATAAATGCTGAACCGTAGTCACCAACTGTTATGCTGGTTGTATCTATTGACTAA
AACTACCCTAGGGATAAATGCTGAAC
CGTGGTTACCAACTGTTATGCTGGTTGTATCTATTAAGTCAACCAACCAATGA
TAAATGCTGAACCATAATTACCAACT
GTTACATTGCTGGTACTACATTAAGAATAAATGCTGCATCTACAAGTACCACCT
GTTGTGTTAATAAATGCTGCACCTGC
TAGTACAACTGTTGCTGGTCATGATAGTTACTACACATTACACACCAGACAGTG
GCAAACAAGGTTATGTAGAAACCA

Fig 3

>113g4 844bp in-house 1-844

ATAGAACTGTTTGATATACAACCTATCTCACTCCCAATTGTGACTTGAATAAATAAATACCTATCACCTAGTAATCTTT
ATCTTAACGTAATCTCTGCAAGCACAATCAATGTATAAAGCATAAAGATAAAATCTTGGTGAGGTTTAAGTTCATAAT
TATAATGAACAACAATTACTAAAAGGATGGTATCAACAATTATAGGCTAGGTAGAACCATAGTGGCTGTTCCGGAGTT
CGGGTAGTTTGGGAAGGTTGGGAAGGTTGGATAGTTTGAGAAGGTTCCGTGGCTGATTCTAAATTAACAGAGAACGATAT
AATGTACAAAAAACATTTCAGAAATTTAAACAACCTTTATATATATATATTAATGCTCTTGTTCATCAACTTGCCATTGC
TGTGATGATGCTTTCTGTTAAATATACCTTTAAGAACAGATTCACTATCTCACTAATATTAACCCCTTATCTTTT
GTTTTGACATTCCATAATGACACAAAAGATTGTGAAATATTTTTAGCCTCAAGGGGATTCTACTCATTCCTCTCAAACA
CACATTCTTTGTATCACAATACCTTTTGTAAACAGAGGAACAAAAATTGACACGCATGTCAATTTACCCCTATAGCACTA
TCACTACAAATCAAAGGATTTACAATAGTGGGAATGTCAAATCATGTATATTATTAACACATTACACATTTTATTTTCA
GGTACATAATACTCAATATCTAAACCTTCAAAATGGTACTGTACCTTAACTTTCTCCTTCATGTCTAGTTGAATATTAT
ACTTGCTAATGTCAAAAAATCATGTCTTCACACATTCCAGGTTGT

Fig 4

>117c_af 623bp in-house: 1-623

AACTGTCCTGTGAAGACGAACATCACAACCACAATCATGGTCATAACCAAAAT
CACAATCATGTTGCTCCTATTCTCTACA
ACAGCTGGACAATCATTAAATAATAAAATTGATACATCTAAAGTGACAGCTCT
CAACATGGCCAACTCTGCTGACGATCT
AGCAAAAGTTTTCAAAGATTCGACTAAAAAATATCAAATCAAACCAATTATCA
AATCAGACAGTGATGAACAAATGATTA
TCAACATTCCATTCTTAATGGTAGTGTCAAATTGTATTTCGATAATTCTACGTAC
CAATGGGGGATTTGTATTGTCCCAA
ACAATAAAATTATTCAAAAATGACACATCAATTGATTTTGATAATGTGGATTTCG
AAGAAACCAATACAGGTGTAACTCA
TCCTCAAGTTGGTGTGTGCTAATAATGATAGCGATGATCTTCCAGAGTTTTTGA
ATCAAATAACGATGACGATTTTGTCTG
AACATTATGTGTCTCGACATAAATCACTGGGGTAAATCAATTGACAATATTTA
TTGAAGATATTTATGATGAANGAGAA
GAAGAGTGTCAATTTACATTCAATTGAATTGAGAAGGGGAATTCAGTGAATTAA
ACAAAGACCC

Fig 5

>15c1 977bp in-house: 1-977 bp

TTTTTTTACAAATATAGTTAGTCTCTTTTTTAAAAATTGAACAACAAAAAAGTAAACTACTATCACCACCACCA
CCACCACCAAAACATCATAGTGGAACTTAATTGAAGAAATATATTAATAACCATTAATTATAATACATACTCAAAAGGAA
TAGGAGTAAAAACCTTTATATGTAATTAATTAAATAGCAAAAAAAGGAAAGATTTCACCAAAATCTTGTGA
ATTAAATTAAATTTTCATTTCTTGAAGTGATATAGTCGTAATAGCAGTAATATTAGCAATATTTAAATAAAC
TTTAAAAATAACAATTATAATATAGTAATAATAACGAATTTAACAAAAAAGGGGGGGAAGACAACGAATAT
AGAAGAAGAAAAACAACAGACGGGTAGTAGATATCTGGCTTAAAAAGCATATCTAAAGTACAGCAACACATAAT
GCAGCAAGACAACCCATTAAACAAGAATCATTACCTCCAGAACGTGGTTGTTGTTGTACATACATAGGTTGTTGTTG
TTGTTGTTGATAATATCCACCACCACCCTGGCTGTTGTTGATAATATCCACCTTGTGGAGGTGGTGGTCCATAAGCAT
TATATCCTTGTGTTGTTGATAGTGGCCATGACCACCACCACCCTAAACATCCCTCGATCTTGTGTTTGTGAGAA
TAATTGGGTTGTGATTGTGGTACATAACTTTGTTGTTGTTGTTGATTGGGTTGATTATTATAATTGGTGGTGACC
ACTAGGTTTACCGAAATATTCGTCTTTGACATTGTTTATATTTATAGTGGTGTAAATTTGGTGTGTTGTTGTTGTTAA
GATTGAGTATATAGAAGTTGGAAAAATTTAATAACAATTAATCTAACTTGATATAAGATGGATTAGCAATGATAATGAAG
AAGTAAAGTTGAATGTG

Fig 6

1 QQSYVPSQP NYSQGTCDRG MFSGGGGGHG HYQQQQGYNA YGPPPPQGGY

51 YQQPPGGGG YVQQQQQQP NYVQQQPRSG GNDSCIMGCL AALCVCTLD

101 MLF

>17g1 731bp in-house:1-604 public:605-731

GCTGTAGTTTTGCTTCCAAAAGTTTGATCTCGTCATCAACATCATTAACTTCATC
TAATAAGGTGAATAATTTGGMTCCM
KGYTCCACGTGSYYGYATCACTTTTWATAGATTTCACTTYGGACA AWACTTTA
TTTCTYYGYGATCCCATTTCTYKGAMA
GATCCGTGTAATGTTTCKGCGYMAGACATGTCTTTATTATATMGTTTCATTTAAA
GAATAGTGACTCTCTGACAACTGATC
AAAGGTCKGTARAATCCTACTTCGTAATTGATATATATGATTATTACCACTCTG
TAGAAACTTGCCAKATTTGACTGAAT
CTTCGTATAACTCTKGTGTGAGCRAKTTTACTCTGTTAGATAAAATACTCGATTG
GTGAKKGTGAAKTGTTGTCMTTGTAC
TGGTACRKGCTGCRGGRARAKKGATRGATTTKATCATCMAACTGTCCATGGT
ATTRKRTAACAGTKCACTTYCTTTGAT
AGAATCAATTAAGTTGTGGTAGTCACTAGATTGGGTTTATGATTGTGCGAGTAA
GTGTGATAGTTGCTCATCACTTATAT
GCTTATCCAGAAATTTATTGTACAGCACCATGTGAGTCTTTTGTAGCTGGTTTA
TATACTTTATTTAANATGAACTCTTC
GGGATCGAGTTCATCTTCATCTTCGGAGGTGGAAGCGGGGATAGAATGTAAAC
GTTTGATAGGGGTGTCTTCTTCTTGAT
AAGGTCCCAGA

Fig 7

.207g4 769bp in-house 1-759

GCAAGATCTAACTCCAGTTTTTTGGTGTAAATGTTACACAAGCAAACAAATATAAATCGAAAAAGCCCCAAATAATTCT
CTTCTACAAATTACGAAAAATGTTTCACATGTATGAAAAAGCTTTATCTATACTATTTCTCCTCCAACCTAGCAGTGAG
AATGATACGTATATCTCTATTAGGATACAGTTATCTATTATTAAGTATAATAATCATGGAGATAAATATATATTAAA
TCGATGGAGTTAACGAGAAAAACAATACAAACCCATTTTGCAGCAAAATGAGACATTTACAGAAAAAACAAGAAAAAG
ACAAATTACTCCATTCAATAATCCCAATAAAAAAATAACAAAGAACCAACGTAATAACAAAAACATCACTAATTTCA
CTTTGAAAAATCTTTACATACTCACTTCTAAAGATTAAATAAAGCGATGCATATTCATCAGAAATTTAGTGTATACAATA
TGCAGGTGATTATGAGCCAGGTGAACAATTCTTTACTAAAACTAGGAGTTGTTTATATACAGTATTTTGTCTAAAC
CTGTCTTAACGTATACAAGATAAGATTGTAAATCGGTTAGAAATAACAAGAGGTGTGGTTGTGGACTTGGTGGTGGTGG
CAAAATTTGAATGATATATTGTTTATCTCAAGTATAGCAAAATACAAAGGSCAAAAGGCTGCACAAAAACAAGAACTTGGATT
GTCCGAATCTCTTACCCCTTTCAGAAATGTCTCGTGTATGTGATCAAT

Fig 8

>222g8 543bp in-house: 143-543 public: 1-142

CTACAAAATGAACAATAGACAAGTTCCTAGTAGAACACCACTAGACACCAACA
CTAATCCATGTACTTTCTCATTGGTAG
TACCACTTTCAAATATCCAATCAAACACATCAAATTTGATTGTACTGTTCTTAT
TTTTGAATGAGCTCAGNTGNNTTCTT
CTGCTACCATAAAAGAATTTGGGAATTTCAAATATTGTACTTTCAAAAGNNGAT
AATGCACTAGTAGAAGTGCCAAATAC

Fig 9

TGTTGTAGTAGGTTTCTTTGTCTTTGGTGTATTAACAGATATTACTTTATTGTCT
TCAATATTGATTTGTTCAAAAGGAT
CCTCAAAATTTGGCGTCTTCGTGCCTAATTCTGAATTGGATAATAGTATTTGAG
CCGTAATCTCATCACTATCGNCTTCT
TTACCATCATCTTCGTACCTATAATAGTAATAATAATNGTCAGACAATTCAGTT
TCACCACTATNAATTTTCAGAATCCGT
TNCGTCAACAAGATTTTTTAAAATAAAATTGTCAAACATTGACGNTGCAGTAG
NGGTTGGAAA

Fig 9 (cont)

>222g9 804bp in-house: 1-575 public: 576-704 PathoSeq: 705-804

TTCCAGAGGCAACAAGCGGAAGAAGCACAACGAAAGAAGGAATTTGAACAAA
AGGCCGAATTTATCATTTNGCATCATTAC
TTGAAATGCGCCGAAGAGAAATAGAGAGGCGGAAACAGCAAAAGGAAAGGG
AACAAAGACAAAAGGAGCACGAAGCAAAG
AGGGATATCAGGATACAACAACCTTTTCAGAGCAGGATTCACGGAGTAATCAAAC
TAAAGAAGAAGAGGAAGTGTTCAAGAA
GGCCCGGTCTACTAATTCGGGAGCAGACGAGACTGGTTTGATGTCAGATAAAG
AGTTTGATGATTCTGCATATTCACCCG
ATTATTTGTTTGAAGAGAATTTGTGGAATAAACCAAATCATCCAGATACAAATC
ATAAAACCAAAAAATATACTGAGAAT
GTGGTTGAAAATCTAGATTCTCCACCAAATGATACATCTGCGTACAATTCAAGT
TTTCATGATGAACTAATATTCAAAA
TGAGATCCAAATACCAGAAAATGACGAGTATGTACCACAGATGAAAGCTACAT
CCAGTGTCAATAATACCACCATCCCTG
CACAAAGAAGACATGAGTCACTTTCCACTTCTGAAAACAAAAGAAGGAAATTT
GAAACAGCCGACGTTGGGGTTGATGGG
TTAGATTCTCCAGTGCGGGCACAACCAGAAAATATCTGCAAAATCCAAGTCTCC
GATAATCCCTGGATGGTATACTTTTTT
ATGGNACCNAAGAGACTGGAAACTCCTGAAGGCCAAATTGCTGTGCAGGGACC
AATAGGTACATTATATTCCTCANGGGG
GNCC

Fig 10

>226c_ef1 766bp in-house 1-766bp
AACGTAATTGTTATATTTTACCAAGGTAACAGGGGACCTCATTATCATTAGTTGTCAATTCAATTACTCCAGAAAACAAGA
AACACAAGACTGTTTGGTGTGCTATTAAGGATAATATATAATCAGGATAAAAGAATTTTTTGGTTAAAGAAAATTA
CAGGGACGGTAAATCATTCTTCTCCCTATAAACCAAAAATCTTATATGTCCCAAGTTAACTTATTAGAATTCGAAGATT
ATTTACTTTACAGTGAATCATTAAACATTTTAATTGAAAGCGAGTTTAGCTCAATGTCTTCAGACACAATGCTTTTCAG
GCACCACCAACAAAAGCACCAGAAAGCTCCATGGATCTGGGTACAATTCCTCAAAAGATCTCCAGCAAGATTGTTTCAAAG
GTGGATATCATCATCATCATCAAAAGATAAGCCAGTATATGCAGAAAAGGCCCTTCTCAAGAAACAAAACATAGCACCGG
AACCAATAAAAATAACTAAACAACAAGTACCAGCTAAACAAAATAGGTACATCTGAAACATCCTCGCCTCTAAGTGTGGCT
TCGAGTCATGATAATTCATGTTCCGATTCAGTGCAGCTTCTATATTTTCTGATTCTAAAAATAACAATAGTATGCAAAAT
GTTACTCACAGATGATATAGAGGACATATTAGAGGACATAGACGATGCTGAGATATACGATGCTGAGAAGGTTACCATAA
CATATATAAGTTCTAAATCATGCTAATACACATTATTAATTATTG

Fig 11

>233c_cp1_full 500bp in-house: 1-500 bp

GAAAAATCAAAACAACAACAACAATTAAGCCAAAGTGATAGTACCAAATCTACTTTAGCAAATGATGAAACAAGAAAAAC
ACTTGATCCTAAAGGCGTTGGAAAGCACTACAACAGGTGATAAAGACACAGTTTCATCAGACAAAGCATCTCTGCCAATTG
AAGATAAAGAAAGTTCACCATCCCTAGCTGGAAGTTCAACATCAACACCAAGTGGAACTGATAAAAAAACATCTCCTAAA
AAATTAGTTACCAATGCTGTCAATAAAGTTGAAAAATAATGATGATTCAAAAAATTCATTAAATGAGGCTGAAAAAGGAGC
TAAAAAATCCAAATCTGGATTGAAAAAATTATTTAACAAGAAGTAGAAGTTGTTTAAATTGTTTCGATATAAATTGTATG
AATTCAGTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT
CAATATAATGTTTATTTTT

Fig 12

>22g3 (5') 535bp in-house: 1-535

AGGTTCCAGTTACCAATTTAGGAAGTGTGTTGCAAGCAGGGCTACCAAATATG
GGTGGCAACACATATGGTAGTAAGTGC
TACCAATGTGGGTGCAAAAAATTTTGCCAAAGTAATTTGTATGGCAATAACAGA
AGTGTGGCGGATTCCNAAGTGAAGGAAT
CTTTGGTGTGTAATAAAAAAAGCAATAGCGACTACGCTACAANAGGCAATCNAT
TATTATTATAAAGTGAAGTTATATAT
ATNTTCTCGGGGGGGGGGGGGGGGNTTNGGNNTCCCCCCCCCCCCCCCCANNTTT
TNTCGGCCCNCCCACCNTNCGGCCTTC
TGGCTCCCCCNCGGGCCNCNNGTAAATNCCTCCACCCNCGGANAANGGNA
AANGGGGAACNANNAAGGGGGGACNNN
NCACCCNATGGGAGGGAAAAATCCCNAAANNTTTNCCCCCNCCCNGCCNAAN
CCNCNTGGGGNGGGCCAAANNCNGGGG
GCTNCNCNCCCTNCCCCCCCCGCCNTNCCCNNTNCCNCGANCTCTNNGNG
GC

Fig 13

>22g3 (3') 426bp in-house: 1-426

CCCCCATATAACGTTGTCAATAGCAATACTCTGTGCGACCCATAGTGTGCACTT
CTCGGTGGTATAAAAAAATTTTTTC
TCCCAAAAAAATCTTCTCCTTTCCACCACTTTTTTCTTCTTCTTCTTCCCCATT
CCCTCCCAATCCCTCATTTTCCC
CATTTCCCCTACCCTCCTGGCCCTGTATTCCAAAATTTTCTCGGGGNTACGCCC
CGAAGANAACCTCCCTCCCACCCACC
CATCTTTGTCNGGNTTCGACCTTCGGCCTCANGGCTCCACCGTCGGGGNTCTTG
TATATTTGTAGACTCCNGGAAAAAGG
GAAAAGGGGAGGAAGAAGGGGGGAAAAAANGGAGGGNGAATCCTT
TTTNTTTTNTCCCCCNCTCTCAAACCNAAA
CCCCNTNTGGGNGGTCNAATTAGGGG

Fig 14

>24gG 522bp in-house: 1-522

GAGCAAGCNTGGCCAAAANNCCCACCTTGTTTTGAATAGGTTTGC GTTGTAYA
GGCAGTYAAATGTGTTTTTTKGGCTTG
ATTTTGAGAAAAAGTTGACTGAAAAACATGCAAGAAACGGGGTGATCATGA
AAATAGTCACACACAAAWWGTCAAAAGA
CNATCACMGAMGACTCAGAATGAGCAGAAGGACTTGTCTGAATTGAGTTTTCA
ATTGTTATTTAGAGTTTTAAATTAGAG
TTGTAAATTTTTGGTCAGATTTTACGAAAAAACTGGCTGAAAAAAAACGAGT
CAGGGTGATTGCMMAAAGAACAGAAACA
ATMSATAATCTTAAATTAAGGTAGTAAAGGCTCTGTGAAGTAATTTAGAGTTTA
AACASGGGGGGGCGAGTCAKKKTTAG
AGTTGTGAGTCTMAAGTTTATTTGGCTAGTGAATTGACTGGCMMGATTGTAA
ACGKGGGTAGRAAAAGACACCCTCMC
SCGTYRMCNTNTTGGCCAATTCAACNCGTCCCNGGANACCGCC

Fig 15

>28gK 475bp in-house: 1-475

CCCCGTTAACCACCTTCTAGGTATACCATTTTCATCTGACTGAATAACTGGTTAGT
CGATTTGTTGTTGAAGAAAAGTGACC
ACCTAGTTTTTTCTGCCAACATTTTTTGCATGAGCCGTCGACGCGTTGTCTTTT
TCTACCCACGTTTAAACAATCTTGC
CAGTCAATTCCCTAGCCAAATAAACTTTAGACTCACAACCTCTAACACTGACTCG
TGCCCCCTGTTTANNCTCTAAATTA
CTTCACAGAGCCTTTACTACCTTAATTTAAGATTATCTATTGTGTCTGTTTGT
GCAANCACCTGACTCGNNNGGNT
TTCAGCCAGTTTTTTTCGTTNAATCTGACCAAAAATTTANGNCNCCGATTTAAAA
CTCTAAATAACNATTTAANGNCGNT
CAAANAGTCCTTCTGCGCGGNGANTCGTCNCTATTGTCTTGNGANATGATGT
GTGTGNCTATNTTCAAGAACAT

Fig 16

>328cl 681bp in-house: 1-681

AACCTAAATATGCCCGATTTAAACAAGTTGATGTATTCACCAATGTCAAATATT
TGGGTAATCCAGTTGCCGTTATTTAT
GATAGTGATAATTTAACCCTCAAGAAATGCAAAAAATTGCTCGATGGACAAA
TTTATCAGAAACAACATTTATATTGAC
TCCAAAATCATCAATTGCTGATTATAGTATTAGAATTTTCACTTCTGGTGGGAA
TGAATTACCATTTGCTGGTCATCCTA
CTTTAGGTACTGCATTTGCATTATTGGAAGATGGTAAAATAAAACCAAATGAC
AATGGACAAATAATTCAAGAATGTGGT
GCTGGATTAGTGAATAATCCGTTGAAAAAACCTAATAATAATAGTAATGA
GTTGCCGTTTTTTGTTATCTTTTGAATT
ACCATATTTCAAATTTTATGAAATTGATGACAAAGTAATCGAGGAATTACAAC
ATTCATGGAATGGAACCAATATTATTG
GTAAACCCGGTACTTATTGATGCTGGTCCAAAATGGGCAGTTTTCCAACCTGGC
TCCGGTAAAGAAGTATTAGACTTGAA

Fig 17

TGGTGATTTAGCACAAAATGAGAGATTAAGTTTAGAAAATGGNTGGACAGGNA
TTGGGNCTTTGGNAACATTATGAAAAT
GGGGATCGGTCCAATTGAGAAAATATTGCTCCTGCTGGTGA

Fig 17(cont)

>33gK_part1 1171bp in-house: 1-588 public: 805-1171 PathoSeq: 589-804

TCTAAAATCTTCAGTNCCATNCAGAGTTTTAGGGAATGTTACAGACTCAACTCC
TTTTGCCATGGGGACATTAGGTTCAA
CATTTTATGCTGTCACCTTCTGTTGKCWGATCTTTCCAAATYTAKGASKKGGYW
ACATTACATTTATTGTTTGTTCCTAA
ACTCAAACCTCCTTCAAGGAATWACATGTTTGGCTGCACACCATCAMTTTGTCT
ATGCATCTKATGGTGATCGTATTGGTA
TTTTTAGACGTGGTAGATTAGAGCATGAATTGGTTTGTGAAGGGAACCTCTACAG
TYAACCAATTATTAGTATTTGGAGAA
TACCTTATTGCTACCACATTAGAAGGTGATATTTTCGTATTTAGAAAACTGAA
GGAAAGAAATTCCCMCTGAATTATA
CMCTACMATCAGAATAATTAWTYCTTTAGTTGAAGGAGAAATTGTGGGATTA
ATTCMTCCACCTACGTATTTAAWWARWR
TWRITGYTSCWMCYACYSWWTCTGTGTTTGTATAAAATGTGAGAACTGGCAA
ATTATTATACAAATCCCGGGAATTACAA
TTCGAAGGCGAAAAAGATTTTCATCAATCGAAGCTGCTCCAGTTTGGATGTAATT
GCTGTTGGTACATCTAATGGAAATGT
ATTTTTATTCAACATTAAGGAGGGAAGTGTGGSCAAAAAATTATTACTT
CTGGAAGTGAATCTTCTTCGAAAGT
TGCCTCGATCTCYTTTAGAACAGATGGAGCACCTCATTTGGTTGCTGGTTTGAA
TAACGGAGACTTATATTTCTACGATT
TAGACAAGAAATCACGTGTTTCATGCTTTGAGAAATGCCCATAAAGAGATTCAT
GGGGGTGTTGCAAACGCCAGATTTTGT
AATGGTCAACCAATAGTATTATCAAATGGTGGTGATAATCATTTGAAAGAATTT
GTTTTTGATCATAATTTAACCAGTTC
GAATTCATCCATTGTTCTCCTCCAAGACATCTCAGATATAGAGGTGGGCATTC
AGCACCACCAGTAGGTATAGAATTTT
GTCAAGAAGATAAAACCCATTTTTATTGAGTGCTTGTAGAGATAAAACATTTT
GGACATTCTACTTTGAGAAAAGATGC
TCAAGCACAGGAATTGTGTCAAAGATTGCAAAAATCTAAGGATGTATAAAG

Fig 18

>33gK_part2 1001bp PathoSeq: 1-1001

ATATCATAACCGCCCAAGGATGAACTTTTGCGAGAACATGGGATTCAAGA
AATAAAAGAGTCGGTAGACATTTGTTA
AACACTATTGATGGTGGCATTGTGAAATCTGTATGTGTGTCTCAGTGTGGTAAT
TTTGGTTTAGTGGGATCATCACTGGG
TGGTATTGGATCATACAACCTTCAAAGTGGATTGTTGCGTAAAAAATATGTTTT
ACATAAACAAGCTGTCACCGGTTTAG
CAATTGATGGAATGAATAGAAAAATGGTTAGTTGTGGTTTAGATGGAATTGTG
GGATTCTATGATTTTGGAAAGTCTGTC
TATTTAGGCAAATTACAACCTGAAGCACCTATAACATCCATGATATATCAAA
ACTGTCTGATCTTGTGCTTGTGCCTT

Fig 18(cont)

GGATGATTTGTCCATAGTTGTTATTGACGTGACTACTCAAAAAGTCATAAGAAT
ATTATATGGTCATACCAACAGAATTT
CAGGAATGGATTTCTCGCCTGATGGGAGATGGATAGTTTCAGTTGCATTGGACT
CCACTTTGCGAACTTGGGACTTGCCA
ACTGGTGGTTGTATTGATGGGGTGATTTTACCAATTGTGGCAACTGCAGTTAAA
TTTTCTCCTATTGGTGATATCTTAGC
GACAACACATGTCTCTGGAAATGGTGTATCCTTATGGACTAATCGTGCCCAGTT
CAAGCCTGTGTCCACCAGACACGTAG
AAGAAGATGAGTTTTCACTATTTTATTACCAAATGCTTCTGGAGATGGCGGTT
CAACAATGCTAGACGGGTTTTTGAC
GAGGATTCTAATGAAGACGGCACTATTGATGAACAGTATACATCTGCTGCTCA
AATTGATGCATCCTTGATTACTTTATC
ATCAGAGCCAAGATCAAAATTCAACACTTTATTGCATTGGATACCATTAAAC
AACAAAGCAAACCGAAAGAAGCACCTA
AAAAACCAGAAAATGCACCTTCTTTTTACAATTGACTGGA

Fig 18 (cont)

>33gK_part3 414bp PathoSeq: 1-414

AAATTGCGTAAATTGGATACAAACGGTAACCACGCATTTGAAAGTGAATTCAC
AAAACATTAAGGGAAGCTGGAGAGAG
TGGACAATTTGAAAGATTTTGACTTACTTAACTTATCTCCTGCTGTATTG
GACTTGGAAATTAGATCACTTAATT
CATTTGTTCCATTGACTGAAATGACAAATTTTATTCAAGCTTTAAATGCTGGTTT
GAAATCAAACGCAAATTATGAAATA
TGGGAAACTTTATATGCCATGTTTTCAACATACATGGTGATGTTATCCATCAG
TTTGAAAATGAACTAGTCTTCATGA
AGCTTTGGAAGAATACAGACAGTTAAATGATGAAAAGAATAACAAAATGGATT
CTTTAGTGAAATATTGTGCTAGTATCG
TAAGTTTTATTAGT

Fig 18 (cont)

>35gK 1334bp in-house: 146-669 public: 1-145 PathoSeq: 670-1334

ACAACGTATAATCGACAGTTTACTATATCTGCTGACTTCAAAACCAATGCATTC
TTCAAGCGTGCTCTGTGATTTCTAT
CATAACATCCACTTTCCGGNGTAATCGGATTACTAAAGCCACAGAATCAAGGT
GAACATCAAGCTTCAACTTCTTTCTTG
GTCCACGAATAATTTAATTTGGTTMTTSKKGSMAMKGCTTTCTACRGTAGGTT
TGAATCTTTCCAACATTGTCTTTGCA
TAGAAACMGCACCAGACAAGAAACATGTCCACTCGACCATCAACYTSKGGGT
AWWGACAAAGTWAATCTGTCTGGATCCT
TTTCATCCAGTTTCCCTGCATKGGAWACAAGTNTGTCCCGCACAGTTAAGACT
GTTTTTATTTTSKTGGTATTAGACTCA
TCAAGTTCCGAAGGAGAGGCATCATTTARGGGWATAGACTCCGCTGAGTTAAT
ACTGGATAAATCACTTATTTTCAGATTC
ACTGACTTGTWCTTCAGTGACCTTATCAAAATCCTCAATGTACTCSGARGCGTW
TTCMCTCMATGTGAAGGCTTTTAAAA
GGGCAACRCTGGTTYCAAAATGCTTTCTTGCRAGTTTGTACKTGACAGAAAAA
TCAAAAACYTTGAAAGATATACCTCTT

Fig 19

CTAAAGTCTTTTAAATCAATTTCTTNTCCTAATTTTTCATCATATAGCTTATGAC
 TTGGCAAACCTCCTTACATACCAT
 ATCCATTACAATGCTAGAAATGTCAATCTTCACTGACGATATAAAGGATGGAA
 GAACTTCAAATAATTTTATAAACTCAG
 GATTGGCTGGTGTATCTGCTGCAGGAGCTCCAGATTTATTGTCCATTTGCTCAC
 TCCATGGACATACATTATTAACGTCC
 ATCTTTTCCATTCTTCAAATTTCTTCGGTGAAATAAATTCGTTGACGRWTTTAA
 AACAGACGTACAATGTGAAAGATAA
 GATCATTAGCAGAGAGCAATTCGAGACTCTTGCTTGAAAGTTTGATTGACACG
 TTTTGTGTGAACATATTGTAGGTGGCT
 AAAAGATTGACTTWRGTAAAATGRAACTTATTAACCCTGGGCCCTCACATTTT
 ACATTTTTCATCTTAAACAAAGKGGTT
 CAAAGKGGAACTTGGTTTGGATCCYTTAWTGGAAWATTTTCYAGKRAATACTT
 TCAAAATCAACTCCAGGAGAGCCACAG
 TGATAATTGAATTGGATTTAGATAAGCGGTTAACTTCCCAATTTTCAGTTTTAC
 CAAACTCTGGTAAATGAAGGTTAAGT
 TTTGTGTCCACCACAACAAGTTTACTAAAAACAGCCTTGAGCATTTTGGAGGCA

Fig 19 (cont)

>36g2 (5') 520bp in-house: 1-520

CGTATAGAGAATAATCCGTTGAAATTGATTGTTCAATCATTATTGTATCTTTTCC
 CTTTTTTTTGTTCTAACCATAATGT
 TAGAATAATTAGAAATTGTCTAAATATATATTCAAGTTTAAACAAAAACAGAAT
 GCTTGCAATAAGATTTGATTTCTAATT
 ACTAATCGTTAATATTTAGTTTGGTGGGGTTTTATTTATCGAAGATGTAGCATT
 ATTTGTATCNAATAGATAAAGAACT
 TGAATTAAATGGCNTAATTTGTTGCAATAGTAAAAAGAAGAAAAGTGGTAAG
 GAGTGAGTGAAAATATTTTTTGTCCCA
 ATTTGAGTNGAAATCTTACACCNAAAAGTTTGGACNAAAAGTTTTTACTAAA
 ATCTGANAATCTNCCTGAATAGAACCG
 ATCATCCNCAATNTCCGATTTTNTGAGGANAGATAGTGGCCCCACCTCNTGGTG
 ATTAGAAGGAGCNCCCATGTTTTACAA
 TATCTATATCCAGAATAACNTGTTTGTGACCTCNCCCCNG

Fig 20(a)

>36g2 (3') 472bp in-house: 1-472

CTCTATATATAGTGAAATATAACATCAAATAATGTACAAAAAGTATAATAAA
 TTGATTTAGAAATGAGAAAAAGAAAAA
 AACTTGAAGTAGTGAAGATATATTTGTTGGCTATCTTTCTTGGTATGGCTCAAT
 TCAGCCAATCTTGGATGAAAGGTTGG
 AGTTTTAGTTTCGTGGTTTATTGATTTGTAAGTACTTTCGGGCTAGAAAGTTNA
 CAAACATGATTAATCTTGATATANAT
 ATTTGTAAACATTTGGTGCTCCNTCTTAATCNCCAAAAAGTTTGGGNCCTA
 TCTTTCCNCCNGAAATCTGTATATGT
 TGANTGANCCGNTCCATTCCTGTTNANTTTNCGANTTTAGTTAAAACCTTTTTG
 TCCCAACCTTTTGGGGTTAGANTTCN
 NCCCCANTGTTGCCNNAATATTNCNCCNCCCTNCCCTTTCCCCNTTTTAC
 NAATGCACCAAGTAAGCG

Fig 20(b)

>38g1 1348bp in-house: 183-940 PathoSeq: 1-182 / 941-1348

TCTCTGGTATAACTTGCACTACCTCATCGCTACCCCGGATTTTTTTTTTGGTATGA
TCTACACGTCCTCATCGCTACCCCA
GATTTTTTTTTCTGGTGCGCCGGACACGCCCTCCGGTCCGCACCGAAAACCGGGG
TAATCTCCGTCGGAGATACACATCCG
CGGACACAAAATCAGATGAGCTACCACCGAAAATTCCGAAATTTCAAAAACCTC
AAAATCCCTAAAAACAACTATCCAGA
NATTATTGCCATGCCCTGAGGATGAGTTTAGTTTTTTAATTTTTTGAAAAATGTC
CAAAACTGGTTGTGCTGTATAGGANG
GGTAAGAATTTGCCATTCTGCCCCTTTGGGTGGGTCAGTCNAAAAAAGANGTA
TCACTCTGGTTTCAACGGGAAACAACN
NAAAATGGGATTAAAMTWATCTCCAGAMCAAACCTTAGCTTMWWACACCCAY
TTTAGTTGTACTSGYGWRCCMAAMMCMAA
TTTTCCATTTTGTGGGGANGGGAATTTARACCAAATTTTTTTTTTTGAAATTT
CGCTMAGTGTYMAGAMCCSCAAAAG
TCACCTTTTTTCGTTTTCMCYACGGCARARGCYCACCGGTTTTKYKTGGKGS
MCRGCCMAATTGAWTTTGTGGGTGSGC
ACGKGGA AAAACAGTTKGTAGTGACACGTTTTTGACGTGTGAAACTGCGCT
CGGAGGTACTATATGCGAAAGCAGAAA
AGACAATTGCAAGAATACAGAGAGTTCTTCTCTGGGCTANNGCAATGTGTTTA
AGGCCAAGTCGACGAGTGGGGAGAGTC
TGGAAGTGATATACACATCACGACCTACTTTATACGCTACGTTCCGGCATGGGC
GAGCCACTGTACGGTGGCAAGCCTGAA
CAGTCCCACACCAGATATCTAACGATTCTGTGTATGGGCACTGATGGGATTTAG
TGGATTACTAGCTGATAGCAAGTATT
GAAAACATAAAACCCGACTCGGGGGTATGCCTTGGCAAGTAGCCGGAGTAAAAT
CTGTGACTTTGCTGAGTGTAACCTCCCT
CCATGGTTGGCGATGTTGACGTGCGCGGCAGTTCTTGTCGTATCACAGTCGCA
CGGACACCACACCGGGAGAATCTTAA
GAGGGCTATATGGATGTGGAACGGTTTGCTTGCTGTGGTAAACACTGGCGGG
CGAGCCGACGTTCCACGGACACAGCAA
TGTGTTTGCAACCAAATAAATAACTTGTACGGTTTGAACGTGTTTTTGGCTGCT
CCTTCCAGTTCTTGCGGGGAGAAGCT
TGGGCGCGGGAAGACCACTACTACGTAGTTATCTGGTTGATCCTGCCAGTAGT
CATATGCTTGTCTCA

Fig 21

>3gG 842bp in-house: 171-842 public: 1-170

ACCANTAAGGGANGAGGTGGACAAATTGAACCAGAATTGGGAGCCGCCAAAT
TGGACGGAAACCCCGGGGAGGCGGCCAA
TTTGGGCGAGAACCACGGATGGACAAGCGGGGCAAGTGCCAAGGAAGTTGC
GTGCCAGTGAAAGTGAATGTTGGAGAGC
TGGCAATGAACTGCCC GGCAAGTGATACCAGGAGATAGGTGTGTATAGATTAT
NATGGAACGCCNATTTTTGCAGTATCA
CGCGTAATAAGGACAGCAGTTGGACATCGGTACATGAGAGAGCAATGTNAGTC
TTGATANTAATGAGCCGTGTTGAAGTA
GTATTTTAATCNAATTTTACTCCCAAAGGACAATGGAGATCTGGAGATAACN
CCACACTAATCGGTTCTAGACATAGAC

Fig 22

TAANCCTGAAAGGGGGTACTACAGCTTGTTTTGAAAAGGTTTGC GTTGTATAG
GCCAGTAAAATGNNNCTTTTNTTNGGG
TAGAATTTGAGAAAAAGTTGACTGAAAAAANC GCAAGANACGGGGTGNTCAT
GAAAATAGACACACACAAAACCGTCAAA
AAACAATGGAAAAGCTTCNNCATACGCAGTAGGAGGTGTCTGAATTGAGTTTG
TATTGTTATTTAGAGTTTTAAATTAGA
GTTGTAAATTTTNGGGTAGAATTTACGAAAAAGTCGAACAAAAAACGACAAG
TCAGGGTGATTGCAAAAAACAGAAAC
AATAGATAATCTTAAATTAAGGTAGTAGAGGCTCTGTGAAGTAATTTAGAGTTT
AAACAGGGGGGCACGAGTCAGTGTTA
GAGTTGTGAAGTTTATTTGGCTNNTGAATTGACTGGCANGAT

Fig 22 (cont)

>480c 731bp in-house: 1-731

TTTGAANTCTTCNCCNCCTGNNNCTNCTCAAAGCGCTCCGCNTTGTGNCTNAAN
GGGCTGGCTCAACGTACTATCAAGNG
TAACATAAAACTANNACTTGGGNCATTACGATAAAGAAACGGNGCCTGGAAT
TCCAGCAATTGNCGCTGGACTTGAGTC
GNAATCTAGGNATTGNTAGNCGAGATTTTNCATATGAANNACCTGGGAAACGG
ATCAACTGGCTTAACANGACCAGTATT
GATGNGGAGAGAAAAGTGGGACTTGCAGANTTTCTCAATANCCTCATTCAAGA
CTCAACACTTCAGAATGAACGAGAAGT
GTTGTGCGNTTTTGCAATTGCCGNCTAATTTTAGATTCACCAAGGATATGTTACA
GAATAATCGAGCNACTTGGATTNTG
NGCAAAATAACTGGTACGATGTATATCGTAAGTTGAAACTGGATATACTCAAC
GAATNGTCTAGCAGCATTAGTGNACAG
ATACATATTCGTGATCGCATTAGNCGGGTCTACCAACCACGGATTCTCGACTAG
GCAGGNCTATTGGTACAGATAAAGAA
GANGCCTAAAGAAGAAGCAGNTGGGTTCCTCAATTCTTTNAGAGNNTAGAAAAT
TTGGTAGTACAGGAAGTTTCCCGATCA
AAGAGGGTGTTGGACCCAAACAGNTTAANNTTTCGGCNGNGNCNTTNTCATTNA
ACAATAAGNACTTCTTNAACACCAAAT
CCNNGTTTTTA

Fig 23

>55g3 1063bp in-house: 533-1014 public: 88-532 PathoSeq: 1-87/1015-1063

TTCNCCCCCATTCCAAGATTCCCCTTGTAAGTAAATTGGTTGGAGAACCNCGTT
GGTTTAATTCCCCCGNCGGGAAANN
ATNGTNGTAAGACAATTCTTTCAACAATTTATGATGTTGCTGCATTTTCGTAA
ATCTGCTAGTAAGAGCACCTAATCT
TCGAAATCAATAGAAAACCTCAATCTTTTCGATTTTCTTGTGGATCACTTTACTC
ATAGGCTCGCTATAAACAAACAAGCT
CACTATGATAACATATAAGGTAAAGTATTGACTTGATGTAGATGTAATCACCA
CAGAAGGCATATCAATACCAAGTCGAA
GAGGAGCCTGTTTCGAATTATCATCATTGTAATTCGCTTGCTCTTCTTGATCCA
GATCTTGTGCATTTGGAGAAAAGCTTG
CTTGACAAAATTTCACTTTTACATAACACATTGTCATAACAGAAAAGCTTTTCAATC
AATAAGTTGTTGGCTCCAGCCCATTT

Fig 24

ACCATTTTGGGTTATTTCTGTTCCCTAGCCATGGTGGCCAATTAGATTTAGCTCC
 ATATGGATTACTAATACTTAACATAT
 CTGGACACCCTACAATATCCTCTTTGTTTAATACAAAAACATTGGCATCTTTTA
 GTAAAAATACCATACTAGTTTCTAAC
 AAGATTTGGTTTGCATTATTCCTGGAATCCCTAATGGACAAAATTTTACCTTTA
 ATAGATGGTGTAGAGATGAGCACNAC
 AGAATCAGGATAATCCTCNCCTTTGTTATTGAATTTGAGGTGGGANCACNTCNAT
 CAAATAATCCTCAACCACTTTTTTCGT
 CNGGTCTGGTCTCNCNTAATATTGTGTTGAAATTGTCCAGTTTCCGTTGTGAAG
 TGANATCNTTGGCCGGTAGANGTCTGT
 GATTTAACCCCTCCCGTGTGNAACGACCGAATANTTCTGACGTCCANNAAAA
 ANCGTCCCTATCGTTGTTTGCNTTTAC
 NCCATCCNCCNTTTTTTCCAGTGTTTTCCATGGACTTGTGTGNCGAGTTATTTTCG
 AAGCTGTGATTTCAATTTCAAAAA
 TGTATGTATTTCAATGTCAAATT

Fig 24 (Cont)

>58gA 724bp in-house: 281-582 public: 209-280/583-724 PathoSeq: 1-208

GTGGTGTTTTGGAAAGTAAGGTGTGATTTGCTTAACAAAGGAAAAGAACGAGA
 CGAGAACCAATTCTCAATATATAGGTC
 TTTCCAGGTAGGAAAACGACCAACTGTGGAAGAATGGCACTACCATTGGTTAC
 CAATACAACACTACTGGTTGTCGTTTCAT
 CTGATAACATACACAACCTACCATGTAGAAGTCATCAATTGTAACTCAGCTAT
 ACTTTATCAATATCAGCGAAATTTAT
 TTAGTTTTCGTTNGAAAATACAGTGAAAAATAAAAAATCTACCCGTCNNCNTGA
 ATTGTNINCCTCTGCANCNACAAATNG
 TNNTTATATTGTGATTCATTCNNAGGCTTGATTNCCANCTATTTNTAAACACT
 ACCTTTCATTTNCTACNTTCNGGGAA
 AATAACNCTTGTTGCTGTTGAAAGACCAATNCCNTTGTGAGTACAGAGGAATA
 CTNCCANTATNCGGCTTATANTTANCT
 AAAATTACAATACATAACAGGGAACCGACNTGTTTNCGTCNTTGATAATGAA
 CAGTTNTGGTNCTNNTGAAAAGTAATC
 CCNAATTTGAATGGNTCGAAAGCAACACAATAAGAGTCTTTGCTTGATATTTG
 CTTCTCCAGAATATAAAATAACGCGTT
 AAAAAGACGTGTTTTCTCTTCAATCGCCATCACAAATAAATTCAACAGAGTAG
 TAGATCCTGTTTTTTTCTTCGCCACA
 TCAC

Fig 25

>60gK 990bp in-house: 445-752 public: 1-140/753-990 PathoSeq: 141-444

ATTACCGATCCGTCGGATTTTAAAACCACAAAATTGCCTGCATTAGCAGAGCT
 AGATATTTTCATAGGGTGCTATATATG
 CAAAGATCTATTGAATGCACCCGTGAGGACACAATGTGATCACACGTACTGTT
 CACAATGTATACGAGAATTTTACTTC
 GAGATAATAGATGTCCGCTTTGTAAAACAGAGGTTTTTGAAAGTGGTCTAAAA
 CGTGATCCATTGTTAGAAGAGATCGTC
 ATTAGTTATGCCTCCCTTAGGCCTCATTGATTACGATTATTGGAGATTGAAAAG
 GTGGAATCGAAGCAAGAGGTAGATCG

Fig 26

TGAGAAATCAGCCAATGAGTCAGCGCTGAATGGTAATAGAAATGTAAACAAC
GATGTTGACGAAACTGTGCGCGTTAAAG
ATCAACTGAATGCAGATAAACTAGGTGAAGAAAAAGGGCAAGCTCAACATGG
GGAACAAGTNAAACGAGCAGACTACTGA
AGTTATTCTGTTGCTATCTGATGATGAAGAGAATGGTTCTGATAGCCTAGTAAA
ATGTCCTATTTGTTTTGAGAGAATGG
AATTAGATGTACTACAGGGAAAGCNTATTGACGACTGTCTAAGTGGAAAGAGC
ACGAAGAGGACGCCTACAGACATTTTA
TCCCCAAAAGCCCAACGACCGAAGCAAATCACCTCCTTTTTCCAACCAACAAT
AGATACCANAACNCCTTCCCCCACCTA
CCAGTTNNGGCGTCNACAACTCCACAGCAACTCCGACAACCTACATTGTTGAA
AGCAAACGTCTCATCTCCATCCCAAGT
GGCGCAAAGTACAGTAAACAAGGGCAAGCCATTACCTAAACTCGATATCAGCA
GCTTGAGTACTAAAAAATAAAAGCCA
AGTTGAGTGATATGAACTACCAACAACAGGTAGTAGGAATGAAATGGAAGC
CAGATACTAGCATTACTATGTGATTTAT
AATGCCAACCTTGACACCAATCATCCTGTA

Fig 26 (cont)

>61gB 602bp in-house: 1-602

ACCTACNNTCACNCNNGNCNCNGCAACACCANCNTNCCNCCNAAANAAAGTC
TTCTTTGAATNAGACNTTTCATCTATTG
GAAGACTTGATCTACCTGAAGGTCTGTTTAAATTAGGATTGAAATGATGGTGT
GTTGTTGTTGTGGGTGGGGATTGTGG
TGATCATGGTTTTGATGCAAGGGTGATCCAGGACCGCGCTGGTTTCTTTCNAAT
TCATCTAAAAGAGCCATATGTCTGGG
TNGCCGTINGTNGTTGTTTCTGAGCAATGAATTCCTGTTGGATTTGCNTGNTGTC
ACGTTCAATNTCTTGATTGAATTTAA
TACCATTTTGAAAAATCATTAAAAATCTGCCACTNCNACAGTTACTCTTCCTTTTGT
CTTANGATACAAATNNTACCCTTTT
NCACGTTTGGCGTTTGTGNCCAATGCAAACNCNGTNCCCCGGGGNGNGGGGNC
CCCCCNGCCNTTGTCCANANCCTGNT
GGNTAGTTATNTTCGCGTTACNATCCCCCCCCACCCNNGCGNGNNCCGCCCA
CCCCCAACGTNCNCCTCCTCTCECCNC
NNCTNGCNGCGCCTNTCNGGGCACCCCTTCNCCNCCCTCNC

Fig 27

>62gB 539bp in-house: 101-539 public: 1-100

ATAATAGAATCTGATTTGAAATAATAGGAACCAACAAGAACAAAAAACA
GAAAAAAGATTTGTATAAACA
ATAGAATCAGAAACAAAAAGCTKTAGTTTGKGAAGAAATTGAAACAATCGGA
AAACAACAATATCAAAGTGMTGCCAAT
AACACTGGTATGTACCTAGATGGATTACCAAGATCTACTACATAAAATAATAR
RGGAGTTCCACTCACTCAAAGAGTTCA
AACCATGGGATAGCAGTGTTTTGTATGAGACGTTACTACGATCAGTATTAATA
CTTTGATCGAACTTTTGGGCMTAGAC
AATCCACCCAGTTWCTWCMCTCACCACCMACMATGATAGTTATAGGTGAA
TTTGAAAATWAAATACTATGGRAATGCA

Fig 28

TTAAGCAAGTCAATCAAYSSWCMKRGCATRKTGCAATWTGCRYWGWAYCAWA
GSATGYAATCSATATTACMRRCYKTKSYY
GRGRYYRKKRATACGCGAWMATWTWKAATCMMMGAGTCYWATYCTGCTGK
WWTCMAAGA

Fig 28 (cont)

>64gB 627bp in-house: 1-627

TNCANCTNCCATNCNCCCAGGCNNNGCCACCCCNCGCNGNCCCCCNTNTTTC
CCCCCTCCTTNGTNGCCCTCNNGGTG
GTGTTTGTGGTGTGACNAATAAANATGGTNTATCATTAGAANAGGACATTGCN
NCGGAAATGACTGTCGACAATAAAGAA
GCAAATATATACAATGGATTATGAANGTGCTAGGATGGATTTGAAAGTTTATC
TGGGTTTATTCCAATGTAAAAATTATT
TGTAATTGATATGGCTAATTATTTTGCTCNATATNTATCACAAAAAATGATTA
AGTTCGAAATGAAATTGGCNTCCATA
TATAAAATTTCTGACAGGAAGAGAAAATTCANGACNTGTTGCCCNAAAAA
AACTTTACCCCNCTCNANTCNTGTNN
GACTTAACCCCCAAAAANAANANNGCTGGCGGCGNAAAAAATAGGAGGGG
GCCGGNNGTTTTTAAAAATTNANNCTT
GAATATGAACCCAANNTTTGNNTTNTTTTTNCCACNCCCCCTTCAAATTTNAT
TCCATGTTCCCAAGANNAGGGNGGNG
GGGGNNGGTTCCNNTTTTAAACCNCCCCCCCCGGGTGGNGGGGNCCGTNTTNT
TTCCGGNGGGGCNT

Fig 29

>65g 441bp in-house: 1-441

TTNCTTATGTAGATGTTGTTTCATGAATTTGTATGAACGGACTATGGCTAGGATT
TGGCCAATCTCGGTATTACTACNTT
TCAAGTTCAAAGATTGGGAACTCGTGTATTTTCGTA CTGTCTACATTTTCTTA
AATTTGATAAACGCATAGTAAGTCTT
TGCTTGATACTATGAGATGATTAGAATTA AAAAGTAGACGACTAGTTTCACT
AGATTTATTGAAGTGTCAAAATATAT
TCAGATTGGTTGCAACTGATGGTCTCGAAAATGCNACAGGATTTTTTTCCCCCA
TTTTTTGCCAATTTTTGTCCNATAGA
GTAGAAAGTACCNGTATNCNAATTGTCCCAAAAAGCGATTATAATCCGTACCA
ATATTTCCAATTTTCNTTTAAACCCTG
TTCNCCTCNGTGTTGTTTGTGAAAACNTAACCANGGTG

Fig 30

>8c_cp 890bp in-house: 287-890 public: 1-124/154-286 PathoSeq: 125-153

ATGCAATTCTCATCCGGTGTCTCTTATCCGCTGTTGCTGGGTCCGCTTTGGCTG
CTTACTCCA ACTCCACTGTTACTGG
CATTCAAACCACTGTGTACCATCACTTCATGTGAAGAAAACAAATGTCACGG
AAACTGGAAGGTTACCACTGGTGTTAC
CACCGTCACTGAAGTTGACACTACGTACACCACCTACTGCCCATTTGTCAACCAC
TGAAGCTCCAGCTCCATCTACTGCTA
CTGATGTTTCTACCACCGTTGTACCATCACCTCATGTGAAGAAGACAAATGTC
ATGAAACCGCTGTCACCACCGGTGTC

Fig 31

ACCACTGTCACTGAAGGTACTACCATCTACACTACCTACTGCCCATTGCCATCT
 ACTGAAGCTCCAGGTCCAGCTCCATC
 TACTGCTGAAGAATCTAAACCAGCTGAATCTTCCCCAGTTCCAACCACCGCTGC
 TGAATCTTCCCCAGCTAAAACTACTG
 CTGCTGAATCTTCCCCAGCTCAAGAAACCACTCCAAAGACCGTTGCTGCTGAAT
 CTTCTTCAGCTGAAACTACTGCTCCA
 GCTGTCTCTACCGCTGAAGCCGGTGCTGCTGCTAACGCTGTCCCAGTTGCTGCT
 GGTTTGGTTGGCTTTGGCTGCTTTGTT
 TTAAGTTTATTAGAGCTTAAATCAAATATTTACAAACAAAATTTTCATTTTCCC
 CCCTTTCCCTTTCTTCATTCTTCAAA
 AAAGGGTTATTTACTATTAATTGATAAAATTTATGGTTTCATGTTAATTTACCCTT
 TTCTTTATAAACATTGGTATTATTA
 TTATCATCATTAGNTTTATTTATATTTTCGTGAGTTTTTCGGNTTTAATTAATTTT
 TTTGGATACATATTAATAAATTTAT
 TTGGTACTAG

Fig 31 (cont)

>80g3 669bp in-house: 1-652 PathoSeq: 653-669

TTGCCAAAATTTTATAAAAAATTGTCAAATTGAAAAGAAGTATTTCCCAAAAT
 AAATTGTTTTTTCATCACAAACCGGTTT
 ATATCGCCATAGNCCATTTTAAATCTTAAAGGTTGATACCAGTTAATTGTTGATTT
 CTCTGTTATAGCCCTGTCTAAATC
 TGTCTATTTCTGGTATCGAATCAAAATGTCGCTCATAATGTGCATGTGCGCAAAG
 ATGTCGTAAAGTTTTGATTTTCACT
 CATCTTAAATTTTTTTTAGTGATTGGCATTTTGTCTTTCACATAGTTTTTATTTT
 TAGTTATCAACCTATCAAATACAC
 CTCCACAACAATGCATCCAAATAATAAAAAATTCATTTAAATCAAAAAAGAAAT
 TTATAGATCGTCGAGAAGCCAAGTCTC
 AAGATATAAAACGTGCATTAACCCATAGGGCTAGATTAAGAAAGAACTATTTT
 AAACATTAGAAAAAGAAGGGTTACAA
 GAGGAGAGGAAGCCTGAAGATGAGAACGATATAAGACCAACCAAGAAGAAG
 GGAATAAATTTTGAAGAACGTGCAGCCAT
 TGTGAAACAACGTAAAGAGGAAAAACGTAAATTCAAACCTAGCAAGTGTACAA
 GCAAAATTGGAAAAGATTGAATCTAATT
 CGAAAGAAAGAGCTTTAAACCGTGACCAC

Fig 32

>85g3 481bp in-house: 1-431

CTAATATACGTCGAGTTCTGGGGGTTGAAAAACGGGTATTTTTGGACCAGCAGAAAAAAGTGGATTTGCGCGTGCA
 CGACCCGAAAAAGGGAGAATTTTGAAAAATGGCGAAATTTGGGGTAAGTTTGAGAGAGTGTGGAGCAACAATAAGA
 GAGGGTGAGCGCAATTTGTAATGGCAGGTCGGCAGGCCAATGAAGATGTGTTGTGCAAAAGATGGAATTTGTAGCGGTTG
 CTGTGCTGAGATATTGGCACTTTTTAAGACCGCATGTTTTGGGTAGCGCTGGGTTAAGACCACATTTTTTTTGTTA
 GAAGACCGCAGAAAAAGAGACACATACAAATCAAGACCGCAGAAAAAGAGACACACATTTAAGACACATTTTTTGGT
 AGCACACACTTTTAAAGACACAGAAAAAGAGCACTTATTTCTAAGACCGCATGTTTGGTAGCACACACTTTTAAAGACCA
 C

Fig 33

>2lg2 667bp in-house: 1-667

TTTAGTTTTATATTGATGATGTTTTTAAGTGCTTGTTTATCATGGTGGATGGAAA
TTAGAATGAGTAAATTGAATGGAAA
ATCACTGCAACACCAACAACAACCACTGGTGGATACGAAAATTTAGTGTACAA
ATTTCTGCCAAAAAATACAATAAAAA
CCGCTTATAGTCTTCTACTGACATAACAACACAAGTCAATAAATCAACAACCTC
ATAAACAATGTAGACTTAATACTATCG
CTTAATTATTTAACTATAATAAATACCCTATAGTATTATGCCTTTGTCAATGTG
TGTAAGAATTTGGTTATTACATATCC
ATGTGTNATATATATGTTGATCAAAAAAACGCGATCTTCTCTTTGGTGTAGTGT
GTTACNCAAAAAATTCANTCTAG
GTCNCTGANAATCACTTGAAAATCAAAAATTTGTTGAAATTGAATTTCTCMA
YTTTGAAATTTTGTGTTGAAATTTTTTT
TTTGCTTTACAAAAGACTCCATTTTGTGTTTCCATTTTACAACCAATTACTTAAT
TCCTCTTTTTCATAATTAATACTA
TCATTACTTACAACACTACAACAACACTACGATCATTTCTAAGAAAAAGCAACGA
GGGCGAATTGAGACATTAATCCCTTT
ATTTTATCATCATGCCTTATACAGAAC

Fig 34

>66g4 579bp in-house: 1-579

CCCCGTTAACCCTTCTAGGGTATACCATTTTCATCTGACTGAATAACTGGTTAG
TCGATTTGTTGTTGAAGAAAAGTGAC
CACCTAGTTTTTCTGCCAACATTTTTCGATGAGCCGTCGACGCGTTGTCTTT
TTCTACCCACGTTTAACAATCTTG
CCAGTCAATTCCTAGCCAAATAAACTTTAGACTCACAACCTCTAACACTGACTC
GTGCCCCCTGTTTAACTCTAAAT
ACTTCACAGAGCCTTTACTACCTTAAATTTARGRTTWTSKAKKGTTTCTGTTTTT
TTGCAAATCACCTGACTYGTTTTT
TTTTCAGCCAGGTTTTTCGTTAAATCTGACCAAAAAATTTACRACTCCTATWT
TTAAAACTCYAAAWWACAATTAAC
TCAATTCAGACAAGTCTTCTGCTCATTCTGAGTCTTCTCTATTGTCTTTTACT
TTTTGTGTGTGACTATTTTCATGAT
CACCCCGTTTCTTGCAATTTTTTTCAGTCAACTTTTTCTCAAAATCAAGCCAAAAA
AACACACCTTTAACTACCTATACAA
CGCAAACCTATTCAAAACA

Fig 35

Sequences with known function, *C. albicans* sequence NOT present in the public domain (ALCES/EMBL)

>CFL (223c_cp) 165bp in-house: 1-165

AACTATTGCCAATGGTAAATATGCCAGTGAAATCGAGAATTTTAATAAGTCGG
TCCCTCTTAAGGTCCCATTCAAATTCA
CTAATGCACAATTGGATCTTTATGCTGCTAGCACACATAACCAAGAGCCAATA
TCCTAGTAACGACGCACCATAGTAGAC
CGAAT

Fig 36

>EF4 (29g3) part1 479bp in-house: 130-479 PathoSeq: 1-129

CGCGAAGNNTCAATCATNTCAGAAGAAATGAAAGAAGGTACTCCGTTCTTTAC
TATTGTGGCAAGAATCCCTGTGATTGA
GGCATTGGGTTTTCCGAGGATATTAGAAAGAAGACATCCGGGGCAGCTAGTC
CTCAATTAGTTTTTGATGGGTATGATA
TGTTAGATATCGATCCATTTTGGGTTCCACATACTGAAGAAGAATTAGAAGAAT
TGGGTGAATTTGCAGAAAGAGAAAAT
GTTGCTAGAAGATATATGAATAATATCAGAAGAAGAAAAGGGTTATTTGTTGA
TGAGAAAGTCGTCNAAAATGCTGAAAA
GCNAAGAACTTTGAAAAGAGATTAGATTATCCNGTTNAACAGGCCATATGTGT
GAAATTGTTCCNAAAAGACAGATACN
ANGTGGNCCNTATTTGTTTAATATCCACNACCAGTTAATGTTTTGATATNGAT
GTTTTATATAGTCCAATGTTGAGAC

Fig 37 (a)

>EF4 (29g3) part2 1706bp

AAGTCATGCGATTGCAACAAGGATCACAAGAACCAGAAGTTCACGAACATTTG
ATTAATTTGATTGATTCACCTGGGCAT
ATTGACTTTTCGTCTGAAGTGAGTACTTCTTCGAGATTATGTGATGGTGCAGTT
GTTTTGGTCGATGTCGTCGAAGGTGT
CTGCTCACAACAGTCAACGTTCTACGCCAATGTTGGATTGATAAGTTGAAGC
CATTACTAGTTATTAACAAAATTGATA
GGTTAATCACAGAATGGAAATTGTCTCCCTTGGAGGCATACCAACACATTTCC
AGAATTATAGAACAAGTAAACTCTGTG
ATTGGGTCATTTTTTGCTGGTGATAGACTAGAAGATGACTTGAATTGGCGTGAG
GCTGGTCTGTGCGGGGAGTTTATCGA
GAAGAGTGATGAAGACTTGTATTTACACACCTGAAAAGAATAATGTAATATTTG
CCTCGGCAATAGATGGATGGGCATTTT
CAGTCAATACATTTGCCAAAATATACCTGAAAAAATTAGGGTTCTCTCAACAA
GCATTGTCAAAAACCTCTCTGGGGAGAC
TTTTACTTGGATATGAAAAATAAAAAATCATCCCTGGTAAAAAATTGAAAA
TAATAGTAACAGTTTGAAGCCATTATT
TGTTTCGTTGATTTTGGACCAGGTTTGGGCTGTTTATGAAAACCTGTGTTATTGA
AAGAAATCAAGACAAGTTGGAAAAAA
TCATTGAGAAATTAGGGGCCAAAATCACCCCTCGTGATTTGCGATCCAAAGAT
TACAAGAACTTGCTAAACTTGATTATG

Fig 37 (b)

TCTCAGTGGATTCTTTGAGTCATGCCATATTGGGGTCAGTGATTGAATACTTG
 CCAAGCCCCATTGTTGCTCAGCGTGA
 AAGAATAGACAARAWWTWRRRKGRMCSMYTATARWRYWKTGKWTTCAR
 AAMTGSATAWWTCCAACTAGTCGAMMCTT
 CATTTGKMAARRMKWTGCASRMAYSYSMKRGTWMAcMCCCRKAAMCCMAT
 WGGCMATWKYAKNTGTMYCAAAATTGTTG
 TCAATCCCCCAATGAAGACTTACCCAAAGCTAGTAATGCCCCGCTACTGGAGGA
 TTGACGGCCGATGAAATCCAAGAACGA
 GGAAGAATTGCTCGAGAATTAGCCAAAAAGGCATCTGAAGCAGCTGCTTTGGC
 ACAAAGAAGGTTCCCAAAAATGAAGAT
 GAGTTTGCCATTAAACCCARGAAAGATCCATTTGAATGGGAATTTGAGGAGGA
 CGATTTTGAGAATGAGGAAGATGAGAG
 CGATGCAAACGCAGTTGAAGAATCAACTGAAACCATAGTGGGTTTCACTCGTA
 TTTATTCTGGATCGTTATCTAGAGGCC
 AAAAGCTCACGGTAATTGGACCCAAATACGACCCTTCATTACNTAGAGACCAT
 CAAACCAACTTTGAACAAATAACCAAGT
 GAAGTGGAAATTAAAGATTTGTTTTATATCATGGGAGGAGAATTAGTGAGAAT
 GGAAAATTTCCGTGCGGGTAATATTGT
 TGGGGTTGTTGGATTGGATAACGCCGTGTTTAAGAATGCCACAATTTGCTCACC
 GTAACGTGAAGATAAACCATACATTA
 ATTTAGTTTCAACATCAACCTTGATCCACAATAAACCAATTATGAAAATAGCA
 GTTGAACCAACAAACCCAATAAACTA
 GCAAAATTGGAACGAGGATTAGATTT

Fig 37(b) (cont)

>NDI (17c_cp) 807bp in-house: 1-614 PathoSeq: 615-807

AACCTATTCCATAATGTTTACTAGATCATTGATTAAAGGTGGTGGCAGACTTGC
 TACTACCAGATCATTGGTCAACAACT
 CTACTAGTTTGGTTTTAAAAATCAATTTAAGAAATATTCAACATCAACTCCTC
 CTAAGGTTGCCAAATCAAAATCTTCG
 ACAATTGGTAAATATTCAGATACACTTTTTACACTGCTGTGATATCGGTTATT
 GGTTCGCGGTTTGATCGGTTACAA
 AATTTACGAAGAGTCTCAACCTGTTGATCAAGTGAACAAACACCATTGTTTCC
 TAATGGTGAAAAAAGAAAACTTTAG
 TTATTTTGGGTTCTGGTTGGGGTGCTATTTTATTGAAAACTTGGATACCA
 CCTTGATAATGTTGNTATTGTCTCC
 CCAAGAACTATTTCTTTTCAACCCATTGTTACCATCTGTTCTACCGGTACTG
 TTGAATTGAGATCTATTATTGAACC
 TGTCAGATCAGTCACCAGAAGATGCCCTGGCAAGTTATTTACCTTGAAGCAGA
 AGCTACAAATATNAACCCCTAAACTA
 ATGAGTTGACACTTAACAAAGTACTACTGTCCGTTCTGGTCATTCTGGTAAAAA
 TACTTCCTCTTCTAAATCAACTGTTG
 CCGAATACACTGGGGTTGAAGAAATCACTACCACCTTGAATTATGACTATTTA
 GTTGTTGGTGTGGTGCTCAACAATN
 CTANTTTTCGGNAATCCTGGGAGNCGCNTGAGGAANTTCAACCCCTTTTTTGA
 AGAANGNCCAGTGGANGCCNTCTGCN
 AATTAGA

Fig 38

>RPL27 (357cL) 560bp in-house: 1-560

AAAAATGGCTAAGTTCATCAAATCTGGTAAAGTTGCTATTGTTGTAAGAGGTC
GTTACGCTGGTAAAAAAGTAGTCATTG
TGAAACCACATGATGAAGGTACCAAATCTCACCCATTCCCACATGCCATTGTC
GCTGGTATTGAAAGAGCTCCATTGAAG
GTTACCAAGAAGATGGATGCTAAAAAAGTTACCAAAGAAGTAAAGTCAAGC
CATTTGTAAATTAGTAACTACAACCA
TTTAATGCCAACTAGATACTCATTGGATGTTGAATCATTCAAATCTGCTGTCAC
TTCTGAAGCTTTAGAAGAACCATCTC
AAAGAGAAGAAGCTAAAAAAGTTGTCAAGAAGGCTTTTGAAGAAAAACATC
AAGCTGGTAAGAACAATGGTTCTTCCA
AAAATTACACTTTTAAGAAAGGAACCACTTTATTTGAATGTTTGAATATAGG
TTGAATCAGAGAGACAAAGTAGAAGA
AAATACAAAAAAGAGAGTATATCTGTATAGTATAATTTAATGGGGGTCTAATT
TACTTACCACCTTTATTCGTGCATTATT

Fig 39

>SADH (110c_af) 650bp in-house: 1-650

AACCTTTTGAAACGATTAANTNCAATCAAACAATCTTATTCAAAAAGTACTCGC
AATACGTACAATGTCAATTCCATCTAC
TCAGTACGGATTTTTTTATAATAAAGCTAGTGGTCTTAATTTGAAAAAAGACTT
GCCGGTTAACAAGCCAGGTGCTGGTC
AATTGCTTTTAAAGGTTGATGCAGTTGGCCTTTGTCATTACAGATTTACATGTTCT
CTATGAAGGTTTGGATTGTGGTGAT
AATTATGTGATGGGCCACGAAATTGCTGGGACTGANGCTGAACTANGGTGAAG
AGGTGAGTGAGTTTGCAGATGGAGATC
GTGTCGCTTGNGTCNGNCCCCANTGGATGTGGNCTTTGCAAACACTGTCTTACT
GGTAACGATAAATGINTGNACCAANT
CGTTTATTGGATTTGNTTTCGGATTGGNTTACAATGGNANGNTNCGANCCATT
TTGGTAGNNANGAGANCNANAACCTTG
GTAAAGATCCNTNTAATGTACNTNCCGAGNAAGCTGCNCTTTTNNNGNNTCCN
TATTGANTCNTACCAANGNTTTAAGG
NATGNAGNAGTGTGNCAACNTGGGAATAATAACACNTTCTTNCCTGATGGN
NTTNTGCNGTTACNCNNTTCAAGNNNN
CNNNAAGCNC

Fig 40

Sequences with known function, partial *C. albicans* sequence present in the public domain (ALCES)

>ABP1 (409c10) 1435bp in-house: 842-1435 public: 1-382/779-841 PathoSeq: 383-778

ATGGAAAAAATTGACATGAATACGTATTCAAACAATATCCAACAAGCATACGA
TAAAGTTGTTAGAGGAGAACCAAATGC
AACATTCGTCGTTTATTCTGTTGACAAAGACGCCACTATGGACGTCACTGAAAC
AGGGGACGGATCATTATACGATTTTG
TTGAACATTTTACTGATGGACAAGTTCAATTTGGTTTACCCAGGGTTACTGTTT
CAGGATCTGACGTCTCCAAGAACATC

Fig 41

TTGTTAGGATGGTGTCTGACAGTGCTCCACCAAAATTGAGATTGTCATATGAC
 AATAATTCTGCTGATGTGTCAAGAAT
 ACTGAGCGGATACCATGTGCAAATTACTGCAATGGATCAAAATGATTTATACG
 TGAATCACTTCTTGAATAGAGNTGGTG
 CTGCTGCTGGTGCAAGATATTCCACTCAAACCTCCGGACTCAAAAAACCATCCC
 CTGCTGCACCTAAACCTACTTCAAAA
 CCTGTTGTTGCTAAATCTAGTTCTGCTTCAAAACCTTCATTTGTACCCAAATCTA
 CTGGGAAGCCTGTTGCTCCAGCTAA
 GCCAAAACCAAAGAACATCACCAAGGATGCTGGTTGGGGTGATGCTGAAGAC
 GTTGAGGAAAGAGACTTTGACAAGAAAC
 CTTTGGATAACGTTCCATCGGCATATAAACCAACAAAGGTTAACATTGACGAA
 TTGAGAAAACAAAAATCAGATACAAC
 AGCTCAACTCCTAAAACATTCAAATCTGAACCACAAGAAGAAAAGAATGACG
 ATGATGGGCCATCCAAACCTTTATCGGA
 AAGGATGAAAGCCTATGATCACGACTCAAGTCGTGATGGAAGATTGACTTCTT
 TACCAAAACCAACGATTGGACATTCTG
 TNGCCGATAAATATAAAGCTAGTGCATCTGGGAATGGTGCTGCTCCTGCGTTTG
 GTGCTAAACCAGCATTGTTGTCACAAT
 CAGTTGATTCAAGAAAGGATAAATTGGTAGGTGGTTTGTGCGAGAGATTTTGGT
 GCTGAAAATGGAAAACTCCGGCACAA
 ATTTGGGCTGAAAAAAGGGGAAAAATACAAAACAGTGGCCTCCGATGAGAAAG
 AAATAACTCAAGTGAAAAAGTTGATGA
 GCCAGAGGAACATCATGCTGCCGACTTGGCCAAAAAATTTGAAGAAAAGGCA
 AATATTGCTGGCGATACTCCTTCCTTGC
 CAACTAGAAACTTACCACCAGCACCACCAGCACGAGAAACCGCAATTCCATCT
 AACGAAAAAGACAAAAAAGAAAAGGAA
 GAGGAAGAACAAGCTTCAGCACCATCTTTGCTACTAGAACTTACCACCACCG
 TNACAAAGACAACCTGAGCCCGACCAG
 AACCAGANGAAGAGGAGGAAGAAGAAGAAGAGGAGGCTCCTGNTTCAAGCTT
 ACCAGCAAGAAATCTCCCCCAC

Fig 41 (cont)

>ADE12 (226c_af2) 993bp in-house: 1-646 public: 647-677 PathoSeq: 678-993

NATAAGGAATGANCCAAAGNAGTGTANNAGNTAATAAGNTNAGANANTTCCA
 AAAGAANNAAAANAACCTAACNACANAN
 NNNTATTANATCCAGTAGAAANACAANATTAAGNNACCANTATCTTNNAAG
 NNTNACGANACNTNGTTCNAATTGTTCA
 TTTGGGGANGNACCCTCAACAACTNTANTNGGGTAAACTTTAACNACACCATA
 GACATTNTNATGGTTNTTGAGGAATA
 CCCAACCCAGTCANAAANACCAACAATACCAGTTGATAATAAGTGANGTATGA
 TAAGTACAGANATCAATATCCAACATT
 AAACGCATTAGCACCTTCANCCAAGATTTTTTTATTGGCAGCNATAGCTTCGTG
 CATNAAGTTGACGGAGTCTGACGACG
 AATGGTCTCAAGGTTTCACGGNATTTTTCAAATCTTGCCAATTCTTCCTTAGGA
 TCATATTCAAATTCACCGTATCTTTT
 TTGTCTACTCTCGACTAATCTCAAATATCTAGTTTTGAATTCTTCCCAAGCTTCT
 GGATCAGGGTTGACTAAATGGTGGA
 CTCTGATACCTGATCTACTTGCCTTGGTTGAGTAAGTTGGACCAATACCTTTAC
 CGGTAGTACCTATTGATTTCTTATTG

Fig 42

GNTGTTAATTCAGCATCTTTCAATTTATCAGCACGTTGATGGAAGTCAAAGACC
 AAATGAGCTCTAGATGAAACAAACAA
 TCTATCAGACAATCTAACCCTTTTGCTTCCAAGTTTTCCAATTCAGCAAAGAA
 GGAAGGAACGTGGATAACAACACCAG
 ATCCAACATAAGTTTTGACATTTAGGATTGACCAAACCAGAAGGTAACATGTGG
 AAGTCATACTTGACTTTACCAACAACA
 ATCGTGTGGCCAGCATTGTTACCACCTTGACATCTGGCACAACATCGATATCA
 TCACATAATAAATCGACTAATTTACC
 TTTACCTTCATCCCCCATTGAGATCCTAATAC

Fig 42 (cont)

>CDC48 2448bp in-house: 95-220/285-1340 public: 1-94/221-284/1341-1373/1783-2273

PathoSeq: 1374-1782/2274-2448

CGTTTCCGAWAGCTGCACCTGAGTTGGCACCTGCTGCTGAACCATTATCAGTG
 GCACCAGCATTTCATTGAATCTAAAG
 CTAGAAAATTGACCTCTTGAGGCTTGCAATTGTTGAGCGTAAGACTCATAACG
 ACGTAATTCAGCGTCTGAAACAGATCT
 TTTTGCGGTCTTCATAGCCTCTTCAAAGTGAGCTCTGGTAATGTAAGGCACAGG
 GTCTTCTTCTTCAACTTCATCTACCT
 TCATATCAACATCTTCAGTTATCACCTGTTCTTTTGCTTCTTTAATCTTGTTAAT
 CTTTACTTGGGCTTCAATAGAGTC
 TTTAATAGCAAATTTAGCAGATCTTTGAACAATATAAGACAAATCTGCACCCG
 AGAAACCGTGAGTGATCTTGGCAATT
 CGTTCAAGTCCAAACCAGGTTCTAATGGAGTGTTTCTCAATTGAGCTTGTA AAA
 TAGACAATCTAGCTGGCTCATCTGGC
 AATGGGACATAAATTAATTGATCCAATCTACCTGGTCTCAATAATGCAGGATC
 AATTTGATCTGGTCTGTTAGTGGCACC
 AATGACAAACACATTCTTCTTAGCATTTCATACCGTCCATTTTCAGTCAACAATTG
 ATTGACCACTCTGTCTGGAGGCACCAC
 CAGCATCACCGTGAGAACCACCTCTAGCTTTGGCAATGGAGTCCAATTCATCC
 AAAAAACACCACAGTAGGAGCAGCAGCT
 CTGGCCTTGTCAAATATATCACGGATATTAGACTCAGATTCACCATAACACATA
 CTCAACAATTCTGGACCTTTGACAGA
 AATGAAATTAGCAGAACTTCAGTAGCAACAGCCTTGGCCAAAAGTGTCTTAC
 CAGTACCTGGTGGACCAAAGAACAAAA
 CACCTTTTGTGGTGCCAATCCGAATTTTGGTATTGATCTGGATGTAAACACG
 GATACTCCACGGTTTCTTTTAATTCA
 TTCTTAATGTTGTCCAAACCACCAATATCATCCCAAGTGACATTAACATTTTCA
 ACAACAGTTTCACGCAAGGCAGATGG
 GTTGGAGTTTCCGAGAGCAAATCTGAAGTTGTCTTGAGTGACACCCAAAGAGT
 TCAACACTTCAGTATCAATGGTTTCTT
 CTTCCAAGTCGATAAGATCCATCTTTTACGGATTGTTGCATAGCAGCTTCTG
 AACATAATGAAGCAATATCAGCACCA
 ACGAAACCATGTGTTTCAGAAGCGATGGCTTCCAAGTCAACATCATCAGCCAA
 TTTCATATTCTTTGTGTGGATTCTCAA
 AATCTCTAAACGTCCTTCAGCATCCGGAACACCAATGTCAACTTCTCTGTGCGAA
 TCTTCCATCAATAGAATTTTGAAATC
 TTCTCAAAGCAGGGTCTGTAGTAGCAGCCAATTAACAACCTACATTAGATCT
 GGCCCTTCATACCATCCATAAGGGGT

Fig 43

TAACAATGGAGAAACAACCTCTTCTTTCTACTTCACCCATTAGTTTTGTCTCTCTT
 TGGGGCAATAGAGTCAATCTCATCA
 ATGAAAATAATGGAAGGAGAATTCTTTTCAGCCTCTTCAAAGCTTTTCTTAAA
 TTGGATTCACTCACCAGCCATTTT
 AGACATAATTTCTGGACCATTTATTAAGAAAAAGAAGGCACCTGTTTCATTGG
 CCACTGCTCTTGCCATAATGGTTTTAC
 CGGTACCAGGTGGACCATACATCAAAATACCCTTTGGTGGCTTAATACCAATC
 GATTTGAATAATTGTGGATGTCTTAAA
 GGCAATTCAACCAATTCTCTAATTTGGGCCATTGTTTCTTACACCTCCCAATAT
 CGTCGTAACCCACTTCATTCAAGCTA
 TTTTCTTCATCTTCACGATTAATAGGTTCTCCTTCACAATGAATAATGGTATCTT
 GAGCAACAATTGCAATTTCTTCAGG
 GTCAACTTCAACAACCTTTGAATTCTACTTGTCTCATACCACCCCTCACAGTGAA
 TAAATCACCTTTTCTCACTGGTCTAT
 AGGCTTCAACAAAATATGGCTTCAAGTAAAGGTGGAATAAGGAACCATTAAATA
 CTTCAACAGTATCAGCAATTGGCAAT
 ACTGAGATTCTGTTGGCATATTTAATATCAGGACATGGATGGACAGTAACGAT
 ATCTCCCAATCTGACACGCAAATTGTT
 ACGAACACATCTGTAACTCTAGCAACGCCATCAGGCATATCATCATCAGCTA
 AAACGATCAACACTGTGTCCTTTCTCT
 TCTTACCTTTACCAAGACTGTATCACCACGGAATAATTGTAACAATTCCATTG
 TGTTTGACGACATGGTTATGACAGAA
 TTGTCATCGTTGGTGGCGTCATCAACGACCAAGGCATTATCTTTCTTTGGTCTTC
 TTAAAATTGCAGTTGCTGTTTTATC
 ATCTACAGCAGAAGCACCAGAAGCATCAAAATGTTGTTTTTTATCTTC

Fig 43 (cont)

>CIT (99g3) part1 1435bp in-house: 803-1435 public: 1-333 PathoSeq: 334-802

TCCAAAACCTTATTGCTTAGCTATACGGTGTAATGGACCAGCTTAACCATTCAAA
 CCAGCAGCTAATGACAAGAATGGAGA
 AGATAAAGCGGTACCAACCAAGTGGGTGGTGTGTGCAGAGACGTTACCACCTT
 CGTGGTCAGAGTGGATGGTAAGGTACA
 ATCTCATTAAATCAACANATTCCTTGGTGTACCAAAACCTAACANACTGGCCA
 AGTTAGCACCGTAATCCAATTGGAG
 TCAATGGCAGCTGGCAATTTACCATCGTGGAAAACGTTTCTGTAAATCTTAGCA
 GCAATGGTTGGCAATTTAGCTAACA
 ATCGATGGAATCTTCGTAAGTGTATTTCCAGTATTCGGATTTGTTGGCACCTTT
 AGCATAAGCTTGGGCAAATTGGGATT
 CAGATTCCAAAGCAGTAACGGCAATGGAGAATTGAGCCATTGGGTGCAAGTGA
 GATGGAGATCTGTCGATCAATTCTTCA
 ACGTGCTTTGGTAATGCTGATCTAGCAGCAAATTCCTCGGATAAAGCCTTAGTT
 TGGGCGTCAGTTGGAACCTTCACCACT
 CAACAACAACCAGAAAAGAGCTTCTGGTAATGGTTCTTCACCACCTGGTGCTTT
 TGGCAATTCCTTTTGAATGTCTGGGA
 TGGTTCTTCTCTGAAACGGATACCTTCAATTGGGTCCAAAACAGAACCTTCCC
 AAATAAACCTTTGATACCTCTCATA
 CCACCGTAAGCTTGTCTAATAAAACTTCACCCAATGACAGTTTTACCGTGTTT
 TTTTTGAATTGTTTAACTTCTTCAG

Fig 44(a)

CTTTGGCTGGCAAGATTCTTCCAATCTTTGTTTTAAGGTCTAAGAAAGCAAGT
TAGTAAATTATCTTTGTATATACGTT
AAAAGTAAATCCCAAGCACATTGCCCGGATCCTCAAAAAAGTGAATACATAC
TGGTTCAGCAGAAGCATATGTTCTGAT
GCTGTTTTTGAAAGTGCTCTTGGCTACGTTGGTTGAACGTTGAATTGATCTGAA
TGCAGACATTGTTTGATAAATATACT
ATATTCTAAAAGAACTTAAAAGATGGAAAAAGGAAAAGAAGAAATGGAAAA
AAAAAAGTAAAAGGAAATCAATTGCAAA
TATATACAAAAAATCCAGCANGAAATCAGTTGAAATTTATATTCCAAATTTTT
GGTTTTAATGGCTCTTGAAGTTGTGG
TGAACAATTTTTTTTTTTTACTTTTTTCTCATGGATTACTTTAGTTTTGGGTC
TGTCGNGCTGCCGTACAACCTTCC
TGNGAAAATTGATTTTTTTTTCTTCTGGNGAGGATTTTTTGGCGTTCTTTGTAA
ACTTCTTATTTGTCCCACTANATGG
AGAGCAAAAAAAAAAAGTTTGACTTTTACTTTTAACCAATCAATTTTCCAGAATN
TGAACGAGAAAAAGGGAAAAAN

Fig 44a(cont)

>CIT (99g3) part2 327bp public: 1-327

CAAAGAGAATTTGCTCTTAAACATATGCCAGACTACGAATTGTTCAAATTGGTT
TCAAACATTTACGAAGTCGCTCCAGG
TGTTTTGACCAACACGGTAAGACCAAGAACCCATGGCCAAATGTGGACTCCC
ACTCTGGTGTCTTGTTACAATACTACG
GTTTGACTGAACAATCTTTCTACACTGTCTTGTTCCGGTGTTTCCAGAGCCTTTGG
TGCTTGCCACAATTGATCTTGGAC
CGTGGTATCGGTATGCCAATTGAAAGACCAAAATCTTTCTCCACTGAAAAATA
CATTGAATTGGTCAAAAACATCAACAA
AGCTTAA

Fig 44s

>HOL1 (409c5) part1 695bp in-house: 98-695 public: 1-97

TTTCTGGATTATCATGTTATTTGGTTAGCTANACGGAATAATGGGATAATGGAA
GCTGAATATCGATTATATNTATTAGT
TATCACTTTAATCATTTAACCCGTAGGGTTAATTATGTTTGGTGTGGTGCCGCT
AGAGAATGGCCATGGCAAGTGATTT
ATGTTGGATTAGGTTTCATTGGGTTTGGTTGGGGATCAATTGGTGATACTTCAA
TGTCTTATTTAATGGATGCTTATCCT
GATATTGTCATTCAAGGAATGGTGGGAGTAAGTATTATTAATAATACTTTGGCT
TGATTTTCACTTTTGCTTGTTCTTA
TTGGTTAGATGGATCAGGAACACAAAACACATATATTGCCTTGTCATTATTGA
TTTTGCTACCATAGCATTGGTTTTCC
CCTTTTTATATTATGGTAAAACATTTAGAAGGAAAACATAAAGACTTTATGTTT
CAATGGTTGAATTGACTCAAGGGATG
GGATAAGAGAGTGAGTGGTAAAAGAATTTTATTAATGATACATTTATTATTAG
AATTACTACTATGGAAATCCGAGTCTG

Fig 45(a)

TGTTTTTTTTAGAAGTATATTTTAGACGTATTTAGAGTTGGTTTTCTCCTTTGTA
CTNTATTTAGCATTTTATAATATAT
TAAATTCAAGTTGCATTAATATATATAAATAAAAAAACCTACNAANAAAAAAN
GN

Fig 45a (cont)

>HOL1 (409c5) part2 762bp PathoSeq: 1-762

GATCAGAATAATGAGGACTTTATACCTGGAACACTCAATATCTATTCTTGGA
GTTGACTCTGAAGATGAAAACGTGAG
TCATTACGATGCTTCCAGTCGACCAAAAGTGAAAACAAAAGGCAATATAATCC
TCTTCCCACAACCATCGAATTCATGCA
ATGATCCATTAAATTGGAGTAAATGGAGAAAGCTAAGTAACTTTTTATTGTCA
TTTTTATTACTGCTTTTACAGCAGCT
ACTTCAAATGACGCTGGATCAATTCAAGATTCACTTAATGAAAAATATGGAAT
TAGTTACGACGCAATGAATACAGGGGC
AGGCGTTTTATTTTGGGTATTGGATGGGGTACTTTCTTTTAAACACCTGCTTCG
TCGTTATATGGTCGAAAAATAACAT
ACTTTATATGTATCTTTCTTGGTTTATTAGGCGCTGTTTGGTTTGCCTTGGTTAA
AAGCACTTCCGACTCAATTTGGTCG
CAATTGTTTGTGGTATTAGTGAGAGTTGTGCTGAAGCTCAAGTACAATTAAGT
TTATCAGAACTTTATTTTGCCATAA
CCTTGGTTCTGTGCTTACGTCCTATATTGTTGCAACTTCCGTAGGTACTTACTTA
GGACCTTTAATTGCAGCCTTTATTG
TTCAAAACATTGGTTTTAGATGGGTTGGTTGGATTGCAGCAATTATTAGTGGTG
CATTATTGTTGTAATTGTTTTTGT
TTAGATGAAACCTATTTTGATCGAGCAAAGTTTACCAAGCCA

Fig 45b

>ESP1 1458bp in-house: 889-1458 PathoSeq: 1-888

CGATTTTCAATTACAAGATATTTTGCATCATGTTGAAAGCAAATGGTTTGGTGG
GTTTATTTTCAAGTATTTTCACTAATG
ACAATGACGTTGAAAATGAATCCAAGAACGTGTTTCATAAATTCAAACAAGAT
TTAATGAAAATTTTGAAAGATTGTTTA
ACCGTAAGTGACGATAAATCGAATATAGAGAGGTTTCTTCAGTTTAATGAATTT
ATTTATTACTGCTTTTACTCAATGGA
GGAATATAATTATGAATTGGTTGATGATTTGATAAAATTTATAACTATAAATAT
GAATTCTCATGGCAGAATAGTTAATT
TTGGCACTAATGTTAAAATTAATANATTACACGAATTAATTAAGAATTTGATTG
ATAAAGTTAATAAAAAACAAACAAAGA
TGTGACTAGCAACAACAAAAACAACAGCAACAACAACAGCAACAACAACAGC
AACAGCAACAATTCCCAACATATTGGTT
TTGATACCTAATGCCAACTGNTCCAATTTCCCAATGGGAAATCGATGGAANTTC
NTTCGTAAGTAAAATCCAATTTCAAG
GAATNGCAATCAANTCCTTANGTTACTTGATCTAGTCAAATCAAACACCAATA
ACAAGAACAAGTTAATGTTTGTGATA
AATCTAATTTGTATTATTTGATTAATCCCAAGTGGTGATTTAATTCGATCAGAAA
ATCGATTCAAAAACTATTTGAATCA

Fig 46

AATCATTTATGGAGAGGGGAAATTGGAAAATTATCAAGTAATGAACATGAAGA
 TTATCAAGATTCAATATTATGTGAAAT
 CTTGAAAAGTCATTTATTTGTTTATATTGGTCATGGTGGTTGTGATCAATATATT
 ANAGTATCAAAAATTATTTAAAAAAT
 TACCTCCTAGTTTATTGTTAGGTTGTTTCATCAGTTAAATTAGATAATTGTAATTA
 TAACTATAATTCCAGTATGTTACAA
 CCACTGGGTAATATTTATAATTGGTTGAACTGTAAATCGTCAATGATACTCGGG
 AATCTATGGGATGTTACTGATAAAGA
 CATTGATATTTTTACACTTTCATTACTACAAAAATGGGGGTAAATAGATGATTA
 TAATGGCAGTGGCCATGATTATGGTA
 TGAAGAAATTGGATTTGACTAATTGTGTTGTTCAAAGTCGAAGTAAATGTACTT
 TGAAATACTTGAATGGATCAGCACCT
 GTGGTTTATGGTCTACCAATGTATTTAAAATAGACATTCTGTTTGCATATAAGT
 TTATATATTTTAATAATAAGAAAAAG
 AGCATAATTTGGATCTTGATTTTGTATTGTTTGGTTTGTATGAACAAATTTTGC
 ACCCAATCACTATCGAACTTTCCTT
 TTTAAACAGAGAACATTTAATCAACATTTATGTTACATTTAAGCGTTTAAATAC
 ATATTTGTGTTAGATAGTTATATAAT
 GTTGTATGCAAACATACA

Fig 46 (cont)

>FAL1 (190g3) 1439bp in-house: 1-770 public: 861-1299 PathoSeq: 771-860/1300-1439

CTTCTTTTAGAGACAATGCAGTGGTTTTCTTACCAGATGCATGACCCCCACCCA
 ATAAAACTATAATCGATCTATTCACA
 GTATTTGATGCCATTTTGATGGTGATGAATGATGTGATGTGATGCTCATCTTAT
 TGGGAGTTTCAAAAAAAAAAAGTTACA
 CTCGAAAAAAAAAAAAATAGCATTATAAATAGAAGCTTTACTATCTTATAGAAC
 AAAACAAAAAACACTATCTTCTAATTA
 ATAATGGATGATTTTGATAGAGATTTAGATAATGAGTTGGAATTTAGTCATAAA
 TCAACGAAAGGAATAAAGGTTTCATCG
 CACTTTTGAAAGTATGAATTTGAAACCTGATCTTTTGAAAGGAATATATGCCTA
 TGGATTTGAAGCACCATCTGCTATT
 AATCTAGGGCTATTATGCAGATCATCAGTGGTAGAGACACAATAGCACAGGCA
 CAATCTGGAACGGTAAACTGCTACT
 TTTTCTATTGGTATGCTTGAGGTTATAGATACTAAATCAAAGAGTGTCAAGCA
 CTTATCTTGTCTCCTACTAGAGAGTT
 GGCAATTCAAATACAAAATGTGGTCATGCATTTAGGAGATTATATGAACATTC
 ACACCCATGCCTGTATTGGTGGGAAAA
 ATGTCGGTGAGGATGTTAAGAAATTGCAGCAAGGGCAACAAATAGTTAGTGGG
 ACACCAGGTAGAGTGATTGATGTGATA
 AAAAGAAGAAATCTACAACTAGAAATATCAAGGTTCTTATTTTAGATGAAGC
 TGATGAACTTTTACAAAAGGGTTTAA
 AGAACAGATCTACGAAATCTACAAACATTTACCACCTTCGGTTCAAGTAGTAG
 TTGTTAGTGCCACATTGCCACGTGAAG
 TATTGGAGATGACAAGTAGTTTACCACTGATCCAGTGAAAATCTTGGTGAAGA
 GGGATGAGATTTTCGCTTCTGGGAATCA
 CACAATATTATGTTCAATGTGAACGTGAAGATTGGAAGTTTGATACACTATGTG
 ATTTGTATGACAACCTTACAATAACT

Fig 47

CAAGCAGTGATATTTTGAATACCAAATTGAAGGTGAATTGGCTTGCTGATCAA
ATGAAAAAGCAAACTTTACTGTTGT
GGCAATGCATGGTGATATGAAACAAGATGAACGAGATTCAATTATGAACGATT
TTAGAAGGGGGAATTCAAGAGTATTA
TATCTACAGATGTTTGGGCAAGAGGTATTGATGTCCAACAAGTCTCGTTGGTAA
TAAATTATGATTTGCCACCGATAAG
GAAACTATATTCATAGAATTGGACGATCAGGTAGATTTGGTAGAAAGGGAAC
AGCTATAAACTTGATACTAAAGATGA
TGTGGTCACTTTAAAAGAATTGGAGAAATATTATTCAACGANAATTAAGGAAA
TGCCAATGAATATTAATGATATAATG

Fig 47 (cont)

>FBP1 (40c_af) 638bp

AACGTTGGCCTGGCCCAGTTAATTCCGTTTCCAAGCAAATGAATGTCGATACCG
ACATCATCACGTTGACCCGTTTTATT
TTACAAGAACAGCAAACCTGTTGCTCCCACCGCCACCGGTGAGTTGTCGTTGTTG
TTGAATGCGCTTCAATTTGCATTCAA
GTTTATTGCCACAATATCAGAAGAGCTGAGTTGGTCAACCTTATTGGTGTTC
TGGCTCTGCCAACTCTACCGGTGATG
TTCAGAAGAAATTGGATGTGATTGGTGATGAGATCTTTATCAATGCCATGAGAT
CTTCCAACAACGTCAAGGTTTTGGTT
TCTGAAGAGCAAGAAGACCTTATTGTGTTCCAGGTGGTGGCACATATGCTGTT
TGTAATGATCCAATTGATGGGTCGTC
CAATATCGATGCTGGTGTCTGTTGGTACGATTTTGGTGTGTACAAGTTGCA
AGAGGGGTCTACTGGTGGCATCAGCG
ATGTCTTGCGTCCTGGTAAGGAGATGGTCGCTGCGGGGTACACCATGTACGGT
GCATCTGCCCATTTGGCATTGACTACA
GGTCACGGNGTCAATCTTTTACTTTGGATCTCANATGGGTGAATTTATTTTGC
CNATCCAACTTGGAAGTTCCAGA

Fig 48

>GAL2 (360c6) 1004bp in-house: 625-1004 PathoSeq: 1-624

TCCATTTTCCCTTTTCTCTTTTCTACATCATCCTCACANCAATTTCAAATATG
TCTCAAGACAACGTCTCATCAACAT
CTACAGCTGAGGCTGTAAATAATGAAATCAAAGTCAAAGATGAATTTCCACAA
GAAGAACAAGCTCATACTAGTTTAGAA
GATAAACCAAGTGAGTGCATACATTGGTATCATCATTATGTGTTTCCTTATTGCC
TTTGGTGGTTTTGTTTTCGGTTTTCGA
TACTGGTACCATTCTGGTTTTATTAATATGTCTGACTTTTTAGAAAGATTCCGGT
GGTACTAAAGCTGACGGTACTCTT
ACTTTTCCAATGTGCAAACTGGTTTAAATGATTGGTTTGTCAACGCTGGTTGTG
CCATTGGTGMWTTATYCTTGTCYAAA
GTCGGTGATATGTATGGTAGAAGAGTTGGTATCATGACTGCTATGATTGYCTAT
ATTGTTGGTATTATTGTTCAAATTGC
TTCTCAACATGCTTGGTATCAAGTCATGATTGGTAGAATTATYACTGGTCTTGC
CGTYGGTATGTTATCAGTTTTATGTC

Fig 49

CTTTGTTCAATTTCCGAGGTTTCTCCAAAACATTTGAGAGGTACTTTGGTGTGCTG
TTTCCAATTGATGATTACCTTGGGT
ATCTTCNTGGGNTATTGGCTACCTATGGTACTAAGAGTTACTCAGACTCTAGAC
AATGGAGAATTCCATTAGGTTTATGT
TTCGCCTGGGCTTTATGTTTGGTTGCTGGTATGGTTAGAATGCCAGAATCTCCA
CGTTACCTTGTTCGGTAAAGACAGAAT
TGAAGATGCTAAAATGTCACCTGCCAAAATAACAAGGTTTCTCCAGAGGACC
CAGCATTATACCGTGAACCTCAATTAA
TCCAAGCTGGTGTTGAAAGAGAAAGATTGGCCGGTAAAGCATCTTGGGGTACT
TTATTCAATGGTAAACCAAGAATCTTT
GAAAGAGTTATTGTTGGTGTCTGTTACAAGCCTTACAACAATT *Fig 49 (cont)*

>KGD2 (98c_cp) 334bp in-house: 139-334 public: 1-138

TTCTAACAACAACATCTTTCTTGGATCTTCAATCAATTCCTTGATGGTTCTTAAG
AAAATAACAGCTTCACGACCGTCAA
CTACTCTGTGGTCGTAAGTCAATGCTAAGTACATCATTGGTCTAGAAACGATTT
GTCCGTTAACAGNAATTGGTCTTTNT
TTAAAANTGTGTAAACCAAATACGGNAGTTTAANGCATTTTTATAATTGGGGT
ACAGTATAATGATCCAATAACACNGNC
ATTANAAATAGTGAAAGAACNCCGGTCATATCTTACAAAGTCAATTTACNAT
TTCTGGCTTTNTTACNCAAATTANANA
TTTCCTTTTNAATA *Fig 50*

>MAA (249c_af) 619bp

AACCCACCTTCAAAGACAAAGAAGATTTTCGTCAAGCAAACGAATGTCAGAGC
AGAAAAGAACCAAGAACTAATCAAATT
TGCCCGTGACAACCTTAACCATTTACCATTACCGAAAAAGACGGAGGTGCAT
GGGAAAACCTATGAACGAATGATCAGTG
GTATGCTCTACAACCTGTTACAAAAAGAATTGGAAACAACACGTATGTCTTGC
AGAGACTACATGTTGGACTACGGCAGT
TTCAGAACTAGAGATTATAAAACAACCCAAGAATTTCTTGATGCAAAATACAA
ACATTTAGAAAGTTTCATTGGACATGT
TGGCAAAAATGCATTTATGGAATATCCAATCTATTTTGATTATGGGTTTAACAC
TTATTTGGGTGATAATTTCTATTCCA
ATTACAATTTGACAATTTTGGATGTTTCCATAGTCAGAATTGGTAATAATGTCA
AGTGTGGTCCCAATGTATCTATCCTT
ACCCCAACACACCCAGTGGATCCCACCTTTCGCGCTATGATCAATTGGAAAATGC
CTTGCTGTGACGGTGGGTAAACGGGGT
CTGGTTGTGTGGAAGCTGTACCATTCTTTGGTGGGGTGACAGTANGTGATGGCA
GCATT *Fig 51*

>MEG1 (55g1) 1380bp in-house: 1-368 public: 497-1096 PathoSeq: 369-496/1097-1380

AATTACAATCTGGTTTGTACTACCATATCCCATTAGTGTTATTGTCATTGTAGA
 TATTGATAATGGTTAAAGGATTGGT
 TTTCATTTTTTGTGTAATGAATGAGCCAAAATAAAAAATCAATTCGATGCGATG
 CAATGAAGTTTAATAAAATTTTTTTT
 TTCTTTATTTCTTTAATCAACCCNNCNATCNTTAAATTGAATCAATACCNACC
 ATTAACATACTTCTATATNCNTATA
 TTINTTTTACAAAATATCNTGGGGNAGANAACAACACTAGTGNTNCNAAAACAAA
 ACNACNTCCTTATCCNTTATTAANA
 NATTTTCNTCCCCAGGNGGGNATTTAAGAACCGTTCCCAGANNATCATCATCAT
 CATCATCACAAAAGAAGAAATCATCAA
 AGAAACAAAGACATAACGATGAAGACGACGAAGAAAATGGTGGCGGTGAAGG
 ATTTTAGATGCTTCTAGTTCAAGAAAG
 ATTTTACAATTGGCAAAAAGAACAACAGGATGAATTGGAACAGGAAGATGAAA
 CACAAAATCACCTTTCATTTGTTCAATC
 ATTAANAATCAACAAATAGATAGTGAAGAAGAAGAAGAGGAAGATGATTAT
 TCAGATTTGGAAGAAGAAGAAGATTG
 AAGAGATATTATATGATGAAGAAGATGCAGAATTTATCCCAAAGATGCAGAA
 TTATTTAATAAATATTTCCAATCCAGC
 GGTGAAGATAATAAATGATGATGATAATTCATTTCAACCAACAATAAATTT
 AGCTGATAAATTTTANAGCCAAAATTCG
 AGAAANAGAATCCCCACAACAACNACAACAGAGTTTTCCAGATAATAGT
 AATGAAGATGCCGTATTGTTACCACCAA
 AAGTCATTTTAGCTTATGAAAAAATTTGGTCAAATTTTATCAACTTATATTCATG
 GGAAATTACCTAAATTATTTAAATTT
 TTACCAAGTTTAAAAAATTTGGCAAGATGTATTATACGTGACAAATCCAAATAG
 TTGGACTCCTCATGCCACATATGAAGC
 AACTAAATTATTTGTGTGCGAATTTATCAAGTAATGAAGCTACAGTTTTTCATTGA
 AACTATCTTGTGTCACGATTCCGTG
 ATTCTATTGAAAATTCGGATGATCATTCAATTAATTATCATATTTATCGAGCATT
 AAAAAATCATTATATAAACCAGGA
 GCTTTTTTCAAAGGGTCTTGTACCTTTAGTCGATGGTTATTGTTCTGTACGTG
 AAGCCACTATTGCTGCTTCAGTGT
 AACTAAAGTTTCTGTCCCTGTTTACATTTCATGTCATTATTGTGGCGTACTGATG
 AATAAAAAACGAGAATCACCTGTAT
 TTGTCCTACGGCGAATATAA

Fig 52

>RNR1 (38) 2562bp in-house: 1-2562

ATGTATGTTTATAAGAGAGATGGCCGTAAAGAGCCAGTACGTTTCGACAAAAT
 CACTGCCAGAGTTCAAAGATTATGTTA
 CGGTTTGAATCCAAACCAGTTGAACCAGTTGCTATTACCCAAAAAGTTATATC
 AGGTGTTTACCAGGGGGTACTACTA
 TTGAGTTGGACAACTTGGCTGCAGAAATTGCTGCTACAATGACAACAATTCAC
 CCAGATTACGCTGTCTTAGCCGCTAGA
 ATTGCCGTATCAAATTTACATAAGCAAACCACCAAACAGTATTCCAAAGTGTC
 TAAGGATTTATATGAATACATTAATCC
 TAAGACTGGGTACACTCTCCTATGATTTCCAAGGAAACCTACGACATCATTAT
 GGAACACGAAGATGAATTAACTCAG

Fig 53

CCATTGTTTACGACAGAGATTTTAACTACAATTATTTTGGGTTCAAGACTTTGG
 AAAGATCATATTTGTTACGTATCAAC
 GGTAAGGTTGCTGAAAGACCACAACATTTGATCATGAGGGTTGCTGTCGGTAT
 TCACGGTAATGATATACCAAGGGTCAT
 TGAAACCTATAACTTGATGTCTCAAAGATTCTTCACCCATGGTTCTCCTTGTTTA
 TTTAACGCTGGTACACCAAGACCAC
 AAATGTCCTCATGTTTCTTGCTTGCTATGAAGGATGATTCTATTGAAGGTATTT
 ACGACACTTTGAAATCGTGTGCTTTG
 ATCTCAAAAAGTGCTGGAGGAATCGGTTTACACATCCACAACATTGTTCTACC
 GGTGCTTACATTGCTGGTACCAATGG
 TACTTCTAATGGTATTATTCCAATGGTAAGAGTATTCAATAACACTGCACGTTA
 TGTCGACCAAGGTGGTAACAAGAGAC
 CTGGTGCCTTTGCCTTGTAAGTACTTAGAACCATGGCACAGTGACATTTTTGATTCA
 TTGATATTAGAAAGAATCACGGTAAA
 GAAGAAATCAGAGCCAGAGATTTGTTCCCAGCTTTGTGGATTCCAGATTGTTC
 ATGAAAAGAGTTGAACAAAATGGTGA
 CTGGACTTTATTCTACCAAATGAGGCCCCAGGCTTGGCTGATGTTTATGGTGA
 CGAATTCGAAGAATTATACACCAAAT
 ACGAAAAAGAAAACCGTGGTAGACAGACCATCAAAGCTCAAAAATTGTGGTA
 TGCTATTTTGGGAGCCCAAACCTGAAACA
 GGTACCCCATTTATGTTATATAAAGATTCATGTAACAACAAATCCAACCAAAA
 GAACTTGGGTATTATCAAATCTTCCAA
 CTTGTGTTGTGAAATTGTTGAATATTCTGCTCCAGATGAAGTTGCTGTTTGTA
 CTTGGCTTCCATTGCCTTGCCATCAT
 TTGTTGAAAATGATGAAAAAGTACTTGGTACAACCTTTGACAAATTACATCAG
 GTCACTAAGGTTGTCACCCGTAACCTG
 AACAGAGTTATTGACCGTAACCATTACCCAGTCCCAGAAGCTGAAAGATCAAA
 CATGAGACACAGACCAATTGCTTTGGG
 TGTTCAAGGTTTGGCTGATGCCTTTATGGAATTGAGATTACCATTTGACTCTCA
 AGAAGCTAGAGAATTGAACATTCAAA
 TTTTGTGAGACTATCTACCATGCTGCTGTTGAAGCTTCAATTGAATTGGCTAAAG
 AAGAAGGTGCCTACGAAACCTATCCA
 GGTCTCCAGCCTCTCAAGGTTTATTACAATTGATTTGTGGAACAGAAAACCA
 ACTGAATTATGGGATTGGGATACATT
 AAAACAAGATTGGCCAAACATGGTATGAGAACTCCTTGTTGGTTGCACCAA
 TGCCTACTGCTTCCACATCACAATTT
 TGGGTAACAATGAATGTTTTGAACCATACTTCTAACATTTACTCTAGAAGAG
 TATTAGCTGGAGAATTCCAAATTGTC
 AATCCATATTTATTGAAGGACTTGGTTGATTTGGGTGTCTGGAACGACGCTATG
 AAAAGTAGTATTATTGCTAACAATGG
 TTCTATCCAAGCCTTACCAAACATCCCTGATGAAATCAAGGCATTGTACAAAA
 CTGTCTGGGAAATCTCACAAAAACATA
 TTATCGACATGGCTGCTGATAGAGCAGCATTTATTGATCAATCTCAATCATTAA
 ACATTACATCAAAGATCCAACAATG
 GGTAAATTAACAGTATGCACTTCTACGGTTGGAAGAAAGGTTTAAAGACTGG
 TATGTACTACTTAAGAACACAAGCTGC
 CAGTGCTGCTATTCAATTTACCATTGATCAAAAAGATTGCTGAGACTGCCGGTCA
 TACGGTTGCAAACTTGGACAAATTAA

Fig 53 (cont)

ACATTAAGAAATATGTTAACAAAGGAAGAGTTGAGAGTGAGAATACCAGTGAT
GCTCCATACAAGTCACCATCAACCGAA
CCAACCTCATTAGAAAGTTCAGTTGCTGATTTGAAAATAAAAGATGAAGGTGA
AAAGCCAGCTGAAGACAAAACCATTTGA
AGAACTCGAAAATGACATTTATAGTGCCAAAGTTATCGCATGTGCTATTGATA
ATCCAGAATCTTGTACAATGTGTTCTG
GT

Fig 53 (cont)

>RPL16 (485cL) 759bp in-house: 1-759

GGAGGTNTCNNTCTCTGATTCTTCTCCCTGCTCCACNCAAGGGCCAACCAACA
ATGAGTCAAGTCGCTCCAAAGTGGTA
CCAATCAGAAAGACGTTCCAGCTNCAAAACAAACCAGAAAAGACTGCTCGTCC
ACAAAAATTACGTGCCTCTTTAGTCCC
AGGTACCGTTTTAATTTTATTGGCCGGTAGATTCAGAGGTAAAAGAGTTGTTTA
CTTGAAGAACTTGGAAGACAACACCT
TATTGGTTTCTGGTCCATTCAAAGTCAATGGTGTTCATTGAGAAGAGTTAACG
CTAGATACGTTATCGCCACCTCCACC
AAAGTCAACGTTTCTGGTGTTGATGTTTCTAAATTCAACGTCGAATACTTTGCT
AGAGAAAAATCTTCTAAATCTAAAAA
ATCCGAAGCTGAATTCTTCAATGAATCTCAACCAAAGAAAGAAATCAAAGCTG
AAAGAGTTGCTGACCAAAAAATCTGTCTG
ATGCTGCTTTATTAAGTGAAATCAAAAAGACCCCATTTATTGAAACAATACTTG
GCCGCTTCATTCTCTTTGAAGAACGGT
GACAGACCACACTTGTTAAAAATTTAATTTAGGTGAAATTAATATTTTGCAAAC
ATGTTTCATGATAAATAACAATGNGGG
CTTTTTAAAGCAATGGGATGGGGATATGGTTAAGAGGGATGGCTTTATATTTTG
AGTTTTTATATATGGGGACCTTTGGT
TTAATAAATGGAANGNTATTGGGCTTCAAATGAACTTN

Fig 54

>RPS21 (328c3) 391bp in-house: 1-391

AACATTAAAGCAAGATGGAAAACGATAAAGGTCAATTAGTTGAATTATACGTC
CCAAGAAAAATGTTCTGCTACCAACAGA
ATCATTAAAGCCAAAGATCACGTTCTGTTCAAATCTCAATTGCTAAAGTTGAT
GAAGACGGTAGAGCTATTGCTGGTGA
AAACATCACTTACGCTTTAAGTGGTTACGTTAGAGGTAGAGGTGAAGCTGATG
ACTCATTAAACAGATTGGCTCAACAAG
ACGGTTTATTGAAGAACGTCTGGTCTTACTCTCGTTAAGAGAATAGAAGAATA
GACAAAATTGATAATTGGGTATTTTAA
GAAATTACTTTTTTATATTGCAAATTAATTTTAATCTTTCTTCTGTGTATATTTA
ATGNCTTAACATAAT

Fig 55

>RVS167 (67g1) part1 733bp in-house: 145-733 public: 1-144

TCTACTTCTGCTTGAGCACTGACCATTTTTCTTCATCTTTAACAGTTCTTTCTTT
CTTCAGTTCATATTTAGAAAAATT
TCTCNTATGACGATCCAAATCCAATTGTTTATGGTCTCTTTTCACTGACATTTTC
CTTATAGCTTGTATAATCTTCAATA
ATTCTTGTGCTGGTTCAACAATTCTTTTTTCAATCAATTCCAAATCGGGTTTTAA
GGTATCTTTGAGATCTTTAACCACT
GCTTGGTACAGTTCCGATGCTTCAATACCTTGTGGGTATCTTCTGGTACCGTA
GCACTGGGGTCCGATAATCTACCACT
GATTGGTTTATAAATCTCAGCCACGGCTTTGGCAAAATCAATTTGTTTCATCTAA
CATCCCATTGACAGCATTGAAATATT
TCTTGGATTCTTCACTCAACTTTTTTGTTCRTTTCGATTCTTTGAATCTTCTT
TCAGCATCGAGATAAACAGCATCT
TGGGTGATTTCTCCCATGTTGAATTTCTGACGCATTGTCTGTGGGGCCCTAAGG
ACACCTTTTTGAATCCTTTAAATGA
CATAAGATTGTAATAATTGAAAAATAAAAGGAGAAGGAAGAAGGAAGGGAG
ATAGTATATGAAAAGGAAGGGGCGGAGG
GAATTAATTGTAGGAAGAAGTGGCATTGCTTTTTGTCGAAAGCATTTTTTGAGC
GTGCGAGAAATTTAATCCAAAAAAT
GTGTGGTGAAAGG

Fig 56(a)

>RVS167 (67g1) part2 1079bp public: 1-523/668-1079 PathoSeq: 524-667

AGTGGTAGATTATCGGACCCCACTGCTACGGTACCAGAAGATACCCTACAAGG
TATTGAAGCATCGGAAGTACCAAGC
AGTGGTTANAGATCCTATAGATACTTTAAACCCCGATTGGAATTGATTGAAAA
AAGAATTGTTGAACCAGCACAGAAT
TATTGAAGATTATACAAGCTATAAGGAAAATGTCAGTGAAAAGAGACCATAAA
CAATTGGATTTGGATCGTCATAAGAGA
AATTTTTCTAAATATGAACTGAAGAAAGAAAGAACTGTAAAGATGAAGAAAA
AATGTTCACTGCTCAAGCAGAAGTAGA
AATTGCTCAACAAGAGTACGATTATTATAATGATTTGTAAAGAATGAATTGCC
AGTTTTGTATCAAATGCAAAGTGATT
TTATCAAACCATTGTATGTATCATTCTATTACATGCAGTTGAATATTTTCTACAC
ATTATACACTAGAATGGAAGAGTTG
AAAATTCCATATTTTATTTTGTCTACTGATATTGTCGATGCNTATACTGCCAAG
AAGGGGAACATTGAGGAACAAACCGA
TTCTATTGGAATCACTCATTTCAAAGTCCGGCATGCCAAATCCAAATTGGAAGC
CACTAAAAGAAGACATGCTGCTATGA
AATAGTCCACCTCCTACNGGTGCCAAGCTCTATGGCATCTACAGGAACTGGTG
GTGAATTACCTGCATACTCCCCAGGAG
GTTACAACCAACCATATGGTGATAGCAAGTATCAACCACCATCTTCTCCAGCA
ACATACCAATCTCCAGTAGTAGCAGCC
ACTGCTCAATCTCCAGCTACTTATCAATCGCCAGTGGCTACTGGACAACCTCCA
TCATATTTACCACAACTCCAGCCAG
TGCTCCACCACCACAAGTTGGTAGTGGCCTTCCAACATGCACGGCTTTATACGA
TTATACTGCACAAGCCCAGGGTGACT

Fig 56(b)

TGACTTTCCTGCAGGAGCTGTTATTGAAATTATACAAAGAACCGAAGATGCC
AACGGATGGTGGACTGGTAAATACAAT
GGTCAAACCGGTGTGTTCCCTGGTAATTATGTGCAATTA

Fig 56(b) (cont)

>SAM2 (36) 1155bp in-house: 1-1155

ATGACTACTTCCAAGGAACTTTCCTTTTCACTTCAGAATCCGTTGGTGAAGGT
CACCCAGATAAGATTTGTGACCAAGT
CTCCGATGCCATTTTAGATGCTTGTTTAGCTGTTGATCCATTGTCAAAAAGTTGCT
TGTGAAACTGCTGCCAAAACCGGTA
TGATTATGGTTTTTGGTGAAATTACCACTAAAGCTCAATTGGATTATCAAAAAA
TCATTAGAGACACCATTAAACACATT
GGTTACGACGATTCTGAAAAAGGTTTTGATTACAAGACTTGTAACGTCTTGGTT
GCAATTGAACAACAATCTCCAGATAT
TGCTCAAGGTTTACATTACGAAAAAGCTTTGGAAGAGTTGGGTGCTGGTGATC
AAGGTATTATGTTTGGTTATGCCACCG
ATGAAACCGATGAAAAATTGCCATTGACCATTTTATTGGCCACAAATTGAAT
GCTGCCTTGGCTTCTGCCAGAAGATCA
GGTTCCTTGCCATGGTTGAGACCAGATACCAAAACCCAAGTCACCATCGAGTA
TGAAAAAGATGGTGGTGCAATTATCCC
AAAAAGAGTCGACACAATTGTTATTTCCACTCAACATGCCGAAGAAATCACCA
CCGAAAATTTGAGAAAAGAAATTATTG
AACATATCATCAAGCAAGTCATCCAGAACATTTATTAGACGACAAAACATC
TACCACATTCAGCCATCAGGCAGATTC
GTCATTGGTGGTCCCCAAGGTGATGCTGGTTTACTGGTAGAAAGATCATTGTT
GACACCTATGGTGGTTGGGGTGACA
TGGTGGTGGTGCCTTCTCAGGCAAGGATTTCTCAAAGTTGATAGGTCTGCTGC
TTATGCCGCTCGGTGGGTGCTAAGT
CGTTGGTGACCGCCGATTGGCCAAAAGGGCCTTGGTGCAGTTCTCCTATGCTA
TTGGGGTTGCTGAACCCACCAGCATT
TATATAGACACCTATGGGACATCTAAATTGAGCACCGAAGCCCTTGTAGAAAT
TATCAAGAATAATTTTGACTTACGCCC
TGCGTAATTGTAAAAGATTAGATTTGGCTCGTCCTATTTATTTTAAAACCGC
TTCTTACGGACATTTTACTAACCAAG
AAAAATTCTTGGGAACAACCAAAAAAATTAATAATT

Fig 57

>SAP (232c_cp) 619bp

AACCTATAATTTTCAGAAAGAGACTAGATTCTGATAGAAATATAGACGCATCA
CTATATTTTGGAATATAGATCCACAA
GTTACGGAGTTGTTAATGTATGAGTTGTTTCATCCAATTTGGTCCCGTCAAATCA
ATCAATATGCCAAAGGATCGTATATT
GAAAACACACCAGGGGTATGGATTTGTGCAATTTAAAACTCAGCAGATGCCA
AATATACTATGGAAATACTACGAGGAA
TAAGACTTTATGGAAGCATTGAAATTGAAACGAATTGATGCCAAGTCTCAG
TCATCAACAAACAACCCAAATAATCAA

Fig 58

ACAATAGGAACATTTGTACAATCAGATTTGATCAATCCAAATTACATAGATGTT
GGAGCTAAACTATTTATCAACAATCT
TAATCCATTGGTCGATGAATCCTTTTTAATGGATACGTTTAGTAAGTTTGAAC
CCTTATAAGAAACCCAATAATTAGAC
GTGATTCAGAGGGACACTCTTTGGGATACGGATTTCTTACGTACGATGACTTTG
AAAGTAGTGATTTATGCATACAAAAA
ATGAACAACACGATTTTGATGAATACCAAAATTGCTATCAGTTATGCATTCAA
NGGATC

Fig 58 (cont)

>SHA3 (83c3) 1376bp in-house: 375-1376 PathoSeq: 1-374

TGNCCTGGAAATCCCCATTACCATTTTAAAGGTACCACCACCCCCCAAANCT
TNGCGACTATCCATCCAGGTATTANC
CCTTGGAGGATTNGCCATAATAATATGGATGGATCATTGGAGCAAGGAGAT
TTGTCCACTAATATCATGGATAGACAA
ATATCCACCAANAATAGTCATAGAAAAGTTCCAAGAACAGATTTTGAANCCCA
ATTATTAATGAAGAATGCCATGTTACA
ATTGATAGAAGCCATTGAATATTGTACGAAAATAATATTTACCATTGTGATTT
AAAACCAGAAAACATTATGGTTAGAT
ATAATCCATACTATGTTTCGTCCAACTATCAATAACAATAATAACAATGGAGAA
GATGATTTATGCTATGCCAACAGTATT
ATTGACTATAATGAATTACACCTCGTGTGATTGATTTTGGTTTAGCTATGGAC
TCTGCTACCATTGTGTTGAATTCATG
TCGTGGATCGTCATTTTACATGGCACCAGAAAGAACCACCAATTATAACACCC
ATCGTTTAATCAACCAATTAATTGATA
TGAATCAATATGAGTCAATTGAAATCAATGGGACAACAGTGACAAAATCAAAC
TGTAATATTTACCTACATTGGCTGGG
GATATTTGGTCATTGGGAGTATTGTTCAATTAATCACTTGTTCAGAAACCCA
TGGCCCATTGCATCATTTGATAATAA
TCAAAATAATGAAGTGTTAAGAATTATATGTTGAATAATAACAAGGCTGTTTT
GAGCAAAATCTTACCCATTTCTCAC
AATTTAATCGCTTATTAGATAGAATTTTCAAATTGAATCCTAATGATAGAATAG
ATTTACCAACTTTATACAAAGAAGTT
ATTCGTTGTGATTTCTTCAAAGATGATCATTACTACTATGCCCAACATCAACAT
CATCACAATCACAATCAAATCAATAA
TGCTTACAATCACTATCAGAAACAACCTAATCAAGCAAGACCTACTGCAAACC
AACCAATTGTATACACCACCGGAAACCA
CCACTTATAATTCATACGCTAGTGATATGGAAGAAGATGAAATTAGTGATGAT
GAGTTTTATTCTGATGAAGAAGATGAA
GATATTGAAGACTATGAAGAGGAAGAGGAAGAGTATTTTGGTAATGAGCAAC
AACACAACAGCAAGTCACAACAGTGAA
TGGTAATTTTGGTCAAGTTAAAGGTACCTGTTATTACGATACCAAAACCAAAA
CAACTACATATATAAAACCACCAGCTG
CATATACTTTAGAGACGCCTAGTCAAAGTGTTGAATACTGTTAAGTTGTACACA
TAAATAATTAATGACAATTAATAATA
ACGATTAATAATATAG

Fig 59

>TPI1 (233c_cp2) 636bp

AACCAATTTTAgAAAC.AATGGCTCGTCAATTTTTCGTAGGTGGTAACTTCAAAG
CTAACGGTACCAAACAACAAATCACT
TCAATCATCGACAACCTTGAACAAGGCTGATTTACCAAAGGATGTCTGAAGTTGT
CATTTGTCCACCCGCCCTTTACCTTGG
TTTAGCTGTTGAGCAAAACAAACAACCAACTGTTGCCATTGGTGCTCAAAATG
TTTTTGACAAGTCATGTGGTGCTTTCA
CTGGTGAAACCTGTGCTTCTCAAATCTTGGATGTTGGTGCCAGCTGGACTTTAA
CTGGTCACAGTGAAAGAAGAACCATT
ATNAAAGAATCCGATGAATTCATTGCTGAAAAACCAAGTTTGCCTTGGACAC
TGGTGTCAAAGTTATTNTATGTATTGG
TGAAACCTTAGAGGAAAGAAAAGGTNGTGTCACTTTGGATGTTTGNGCCAGAC
AATTGGGATGCTGGTTCCAAGATTGNN
TTGATTGGTCAAACATTGNTGGNNCTTACGAACCTGTTTTGGNCAATTGGGTCT
GGTTTANCCCGNTNCCCCANAAGATG
CTGAAGAAACCTACAAGGTNTTAGACTCATTGGNCAAGANCATTNGTGNCNA
ACAAC TGAAAAACCGANTNTNG

Fig 60

>135g 859bp in-house:1-859

CGTGCATAATTATCTTAAACCGTAGATAAGCAAAAATTTATCTTATGAAATGTTTCAGCGATAAGAGAAAGAAATCAG
GTACCACGAGGAGTGTTTTGAAGAAAACAACTCGTAAATTAATGAATCTAGTTTCTCTACTTGAATAATTTTGAGT
TTTCTGGAAGAACACCTGTTCAGTTTCAAATTAACAAGAAATGTGAAAGAAATAAAATTTGATTTATTCTAGCCTGTT
AATAATCCAGGAAACTCAATTTTCGTAAATGGCAACTTGTCCGAGTGGTTAAGGAGAAAGATTAGAAATCTTTGGGCT
TTGCCCGCGCAGGTTGAGTCCGAGTTGTCGTTATTTTTTTGGTTTACTCTCTATTTTAAATTTAAACTAATCAA
CTGAAACTGGAGTACCTGCCATGATATGAGTAAATACTTTTTTGATTTAAAAATCTATATAAACTCCCTATTTATTTT
TTAATTTAAACCCAGATATTGTCCCAATAATAGTTTTTTGTTTGAACCTATTGCTTTGATGAACTTGTAGTTTAAATC
TTTCCAATTTCACTCTCTTAGTTGGCCACATCAAGTGGCTCATTGAATAATTCTGATCTTGAAGTGTACCAGATGATT
CTGACAAAACCTGCACACGGACCCAGTCAATAGCATTATAGATATTTTGAATTTAAAGTTACCGAATATATCGAATATCTT
TATTGGCCATCTCATCTCTCTTCTGCAATAAATCTTAAACGCTACTTTTCTCAACCTTATTATCCCTCTAGATAC
TCTTCCAAATCTTCAGGTTCAAATATCACTTTAACCATCAATGAACAACCTAGGGCAAAAC

Fig 61

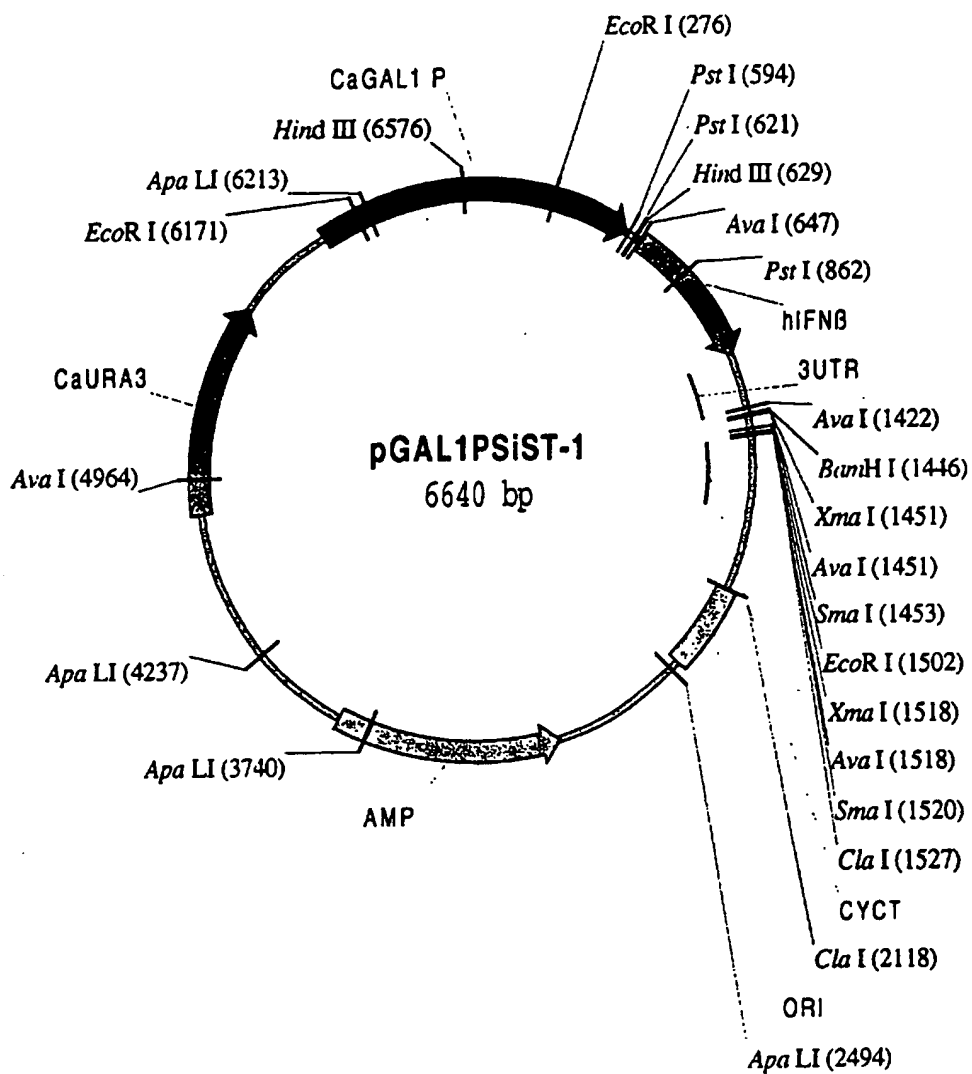


Fig 62

1 TTCCATCGGG GAAAGTGGGG GGGAAAAAAT TTAAAGCAGT TCACAAAACC
AAGGTAGCCC CTTTCACCCC CCCTTTTSTA AAATTCGTCA AGTGTTTTGG
.....
51 TTCCAAAAA TATATGGACA AAGATGATTG TATTTTCCCG ACACCAAAT
AAGGTTTTTT ATATACCTGT TTCTACTAAC ATAAAAGGGC TGTGGTTTAA
.....
101 CATAATTAAT TATGAGAAAG TTAAATGTAA CGTTACAATT TATGTTTATT
GTATTAATTA ATACTCTTTC AATTTACATT GCAATGTAA ATACAAATAA
.....
151 TGAAGGTGAA AAGCGATTTA TGATTTTCC GAAATGAAAA TTTTTTTATG
ACTTCCACTT TTCGCTAAAT ACTAAAAAGG CTTTACTTTT AAAAAAATC
.....
201 GTTTATTTTT TTTGTCGGGC AAAGAAAAAC TGAACAAGGA TTATTAAAT
CAAATAAAAA AAACAGCCCG TTTCTTTTGG ACTGTTCCT AATAATTTTA
.....

EcoRI

251 TTTTGGTGTG TGTGTGTGTC TGGAGAATTC ATTCTCTCTC CATCTTCACA
AAAACCACAA ACAACACAG ACCTCTTAAG TAAGGAGAGA GTAGAAGTGT
.....
301 CAATGTTTAG ACATCTGACA CGATTCAATG TAGTTCGGTT TCCGGGGTTG
GTTACAAATC TGTAGACTGT GCTAAGTACT ATCAAGCCAA AGGCCCAAC
.....
351 GTGTTTAGTT TTCGTTTTTC TTTTTTTTTG GAAAGAATGT TTTAGCTCAT
CACAAATCAA AAGCAAAAAG AAAAAAAAC CTTTCTTACA AAATCGAGTA
.....
401 TGGTTTCTT TCTTCATTCA ATAGTTTGA AAGAATTGTC CCACTTGTTA
ACCAAAAGAA AGAAGTAAGT TATCAAACT TTCTTAAACG GGTGAACAAT
.....
451 TTACAATCAT ATAAATTAA ACTTTGATAT AAAATAGAGT TTGAAAGTTT
AATGTTAGTA TATTTTAATT TGAACTATA TTTTATCTCA AACTTTCAAA
.....
501 CCCAGATCCT TTTGATTTC TTTGTAAAT TTTTCTTCTC CCACATATAC
GGGTCTAGGA AAACTAAAG AAACATTTAA AAAAAAGAG GGTGTATATG
.....

PstI

551 ACACATACAA ACCGATTTTT ATAAGAAAGA GTTATACCTT GCAGCTCGAC
TGTGTATGTT TGGCTAAAAA TATCTTTCT CAATATGGGA CGTCGAGCTG
.....

PstI

HindIII

AvaI

601 CTCGACTGTT TAAACCTGCA GGCATGCAAG CTTGGCCAAA AAGGCCTCGA
GAGCTGACAA ATTTGGACGT CCGTACGTTT GAACCGGTTT TTCCGGAGCT
.....

AvaI

-
651 GGAACATGAC CAACAAGTGT CTCTCCAAA TTGCTCTCCT GTTGTGCTTC
CCTTGTAAGT GTTGTTCACA GAGGAGGTTT AACGAGAGGA CAACACGAAG
.....
701 TCCACTACAG CTCTTTCCAT GAGCTACAAC TTGCTTGGAT TCCTACAAAG
AGGTGATGTC GAGAAAGGTA CTCGATGTTG AACGAACCTA AGGATGTTTC
.....
751 AAGCAGCAAT TTTCAAGTGC AGAAGCTCCT GTGGCAATTG AATGGGAGGC
TTCGTCGTAA AAAGTCACAG TCTTCGAGGA CACCGTTAAC TTACCTCCG
.....
801 TTGAATACTG CCTCAAGGAC AGGATGAACT TTGACATCCC TGAGGAGATT
AACTTATGAC GGAGTTCCTG TCCTACTTGA AACTGTAGGG ACTCCTCTAA
.....

Fig

63

PstI

851 AAGCAGCTGC AGCAGTTCCA GAAGGAGGAC GCCGCATTGA CCATCTATGA
TTCGTCGACG TCGTCAAGGT CTCCTCCTG CGGCGTAACT GGTAGATACT
.....
901 GATGCTCCAG AACATCTTTG CTATTTTCAG ACAAGATTCA TCTAGCACTG
CTACGAGGTC TTGTAGAAAC GATAAAAGTC TGTCTAAGT AGATCGTGAC
.....
951 GCTGGAATGA GACTATGTGT GAGAACCTCC TGGCTAATGT CTATCATCAG
CGACCTTACT CTGATAACAA CTCCTGGAGG ACCGATTACA GATAGTAGTC
.....
1001 ATAAACCATC TGAAGACAGT CCTGGAAGAA AAAGTGGAGA AAGAAGATTT
TATTGGTAG ACTTCTGTCA GGACCTTCTT TTTGACCTCT TTCTTCTAAA
.....
1051 CACCAGGGGA AAAGTCATGA GCAGTCTGCA CCTGAAAAGA TATTATGGGA
GTGGTCCCTT TTGTAGTACT CGTCAGACGT GGACTTTTCT ATAATACCCT
.....
1101 GGATTCTGCA TTACCTGAAG GCCAAGGAGT ACAGTCACTG TGCCTGGACC
CCTAAGACGT AATGGACTTC CGGTCCTCA TGTCAGTGAC ACGGACCTGG
.....
1151 ATAGTCAGAG TGGAAATCCT AAGGAACTTT TACTTCATTA ACAGACTTAC
TATCAGTCTC ACCTTTAGGA TTCTTGAAA ATGAAGTAAT TGTCTGAATG
.....
1201 AGGTTACCTC CGAAACTGAA GATCTCCTAG CCTGTGCCTC TGGGACTGGA
TCCAATGGAG GCTTTGACTT CTAGAGGATC GGACACGGAG ACCCTGACCT
.....
1251 CAATTGCTTC AAGCATTCTT CAACCAGCAG ATGCTGTTTA AGTGACTGAT
GTTAACGAAG TCGTAAGAA GTTGGTCGTC TACGACAAAT TCACTGACTA
.....
1301 GGCTAATGTA CTGCATATGA AAGGACACTA GAAGATTTTG AAATTTTTAT
CCGATTACAT GACGTATACT TTCCTGTGAT CTCTAAAAC TTAAAAATA
.....
1351 TAAATTATGA GTTATTTTTA TTTATTTAAA TTTTATTTTG GAAAATAAAT
ATTTAATACT CAATAAAAAT AAATAAATTT AAAATAAAAC CTTTATTTTA
.....

XmaI

SmaI

BamHI

AvaI

AvaI

1401 TATTTTGGT GCAAAAGTCC CTCGAGGCCT AGCGGCCGCC TAGAGGATCC
ATAAAAACCA CGTTTTCAGG GAGCTCCGGA TCGCCGGCGG ATCTCCTAGG
.....

XmaI

SmaI

AvaI

1451 CCGGGCGCTA GCGCGCCGCT AGGCCTTTTT GGCCAAGCTC GAATTTTCGAG
GGCCCCGAT CCGCCGGCGA TCCGAAAAA CCGGTCGAG CTTAAAGCTC
.....

XmaI

SmaI

EcoRI

AvaI

ClaI

1501 GAATTCGAGC TCGGTACCG GGGGATCGAT CCGTCCCCCT TTTCTTTGT
CTTAAGCTCG AGCCATGGGC CCCCTAGCTA GGCAGGGGGA AAAGGAAACA
.....

1551 CGATATCATG TAATTAGTTA TGTCACGCTT ACATTCACGC CCTCCCCCA
GCTATAGTAC ATTAATCAAT ACAGTGCGAA TGTAAGTGCG GGAGGGGGGT

1601 CATCCGCTCT AACCGAAAAG GAAGGAGTTA GACAACCTGA AGTCTAGGTC
GTAGGCGAGA TTGGCTTTTC CTCCTCAAT CTGTTGGACT TCAGATCCAG

1651 CCTATTTATT TTTTATAGT TATGTTAGTA TTAAGAACGT TATTTATATT
GGATAAATAA AAAAATATCA ATACAATCAT AATTCCTGCA ATAAATATAA

1701 TCAAATTTTT CTTTTTTTC TGTCACAGCG CGTGTACGCA TGTAACATTA
AGTTTAAAAA GAAAAAAG ACATGTCTGC GCACATGCGT ACATTGTAAT

1751 TACTGAAAC CTGCTTGAG AAGGTTTGG GACGCTCGAA GGCTTTAATT
ATGACTTTTG GAACGAACTC TTCCAAAACC CTGCGAGCTT CCGAAATTAA

1801 TGCAAGCTAG CTGGCGTAA TCATGGTCAT AGCTGTTTCC TGTGTGAAAT
ACGTTTCGATC GAACCGCATT AGTACCAGTA TCGACAAAGG ACACACTTTA

1851 TGTATCCGC TCACAATTCC ACACAACATA CGAGCCGGAA GCATAAAGTG
ACAATAGGCG AGTGTTAAGG TGTGTTGTAT GCTCGGCTT CGTATTTTAC

1901 TAAAGCCTGG GGTGCCTAAT GAGTGAGCTA ACTCACATTA ATTGCGTTGC
ATTTCGGACC CCACGGATTA CTCACTCGAT TGAGTGTAAT TAACGCAACG

1951 GCTCACTGCC CGCTTCCAG TCGGAAACC TGTCGTGCCA GAGATCTCTG
CGAGTGACGG GCGAAAGGTC AGCCCTTTGG ACAGCACGGT CTCTAGAGAC

2001 CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC GTATTGGGCG
GTAATTACTT AGCCGGTTGC GCGCCCTCT CCGCCAAACG CATAACCCGC

2051 CTCTTCCGCT TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTTCGGCTGC
GAGAAGGCGA AGGAGCGAGT GACTGAGCGA CCGGAGCCAG CAAGCCGACG

ClaI

2101 GCGAGCGGT ATCAGATCGA TCTCACTCAA AGGCGGTAAT ACGGTTATCC
CCGCTCGCCA TAGTCTAGCT AGAGTGAGTT TCCGCCATTA TGCCAATAGG

2151 ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA
TGTCCTAGTC CCCTATTGCG TCCTTCTTG TACACTCGTT TTCCGGTCGT

2201 AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT TTCCATAGCG
TTTCCGGTCC TTGGCATTTT TCCGGCGCAA CGACCGCAA AAGGTATCCG

2251 TCCGCCCCC TGACGAGCAT CACAAAAATC GACGCTCAAG TCAGAGGTGG
AGGCGGGGG ACTGCTCGTA GTGTTTTTAG CTGCGAGTTC AGTCTCCACC

2301 CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC CTGGAAGCTC
GCTTTGGGCT GTCCTGATAT TTCTATGGTC CGCAAAGGGG GACCTTCGAG

2351 CCTCGTGCGC TCTCTGTTT CGACCCTGCC GCTTACCGGA TACCTGTCCG
GGAGCACGCG AGAGGACAAG GCTGGGACGG CGAATGGCCT ATGGACAGGC

2401 CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCATAGCTC ACGCTGTAGG
GGAAAGAGGG AAGCCCTTCG CACCGCGAAA GAGTATCGAG TGCGACATCC

ApaLI

2451 TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT GTGTGCACGA
ATAGAGTCAA GCCACATCCA GCAAGCGAGG TTCGACCCGA CACACGTGCT

2501 ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG
TGGGGGGCAA GTCGGGCTGG CGACGCGGAA TAGGCCATTG ATAGCAGAAC
.....
2551 AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC AGCCACTGGT
TCAGGTTGGG CCATTCTGTG CTGAATAGCG GTGACCGTCG TCGGTGACCA
.....
2601 AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG AGTTCTTGAA
TTGTCTTAAT CGTCTCGCTC CATACTCCG CCACGATGTC TCAAGAACTT
.....
2651 GTGGTGGCCT AACTACGGCT AACTAGAAG GACAGTATTT GGTATCTGCG
CACCACCGGA TTGATGCCGA TGTGATCTTC CTGTCATAAA CCATAGACGC
.....
2701 CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG CTCTTGATCC
GAGACGACTT CGGTCAATGG AAGCCTTTTT CTCAACCATC GAGAACTAGG
.....
2751 GGCAACAAA CCACCGCTGG TAGCGGTGGT TTTTGTGTT GCAAGCAGCA
CCGTTTGTGTT GGTGGCGACC ATCGCCACCA AAAAAACAA CGTTCGTCTG
.....
2801 GATTACGCGC AGAAAAAAG GATCTCAAGA AGATCCTTG ATCTTTTCTA
CTAATGCGCG TCTTTTTTTC CTAGAGTTCT TCTAGGAAAC TAGAAAAGAT
.....
2851 CGGGGTCTGA CGCTCAGTGG AACGAAAACT CACGTTAAGG GATTTTGGTC
GCCCCAGACT GCGAGTCACC TTGCTTTTGA GTGCAATTCC CTAAACCAGG
.....
2901 ATGAGATTAT CAAAAAGGAT CTCACCTAG ATCTTTTAA ATTAAAAATG
TACTCTAATA GTTTTCTCTA GAAGTGGATC TAGGAAAATT TAATTTTAC
.....
2951 AAGTTTAA TCAATCTAAA GTATATATGA GTAACTTGG TCTGACAGTT
TTCAAATTT AGTTAGATT CATATATACT CATTGAACC AGACTGTCAA
.....
3001 ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG TCTATTTCTG
TGGTTACGAA TTAGTCACTC CGTGGATAGA GTCGCTAGAC AGATAAAGCA
.....
3051 TCATCCATAG TTGCTGACT CCCCGTCGTG TAGATAACTA CGATACGGGA
AGTAGGTATC AACGGACTGA GGGGCAGCAC ATCTATTGAT GCTATGCCCT
.....
3101 GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA GACCCACGCT
CCCGAATGGT AGACCGGGG CACGACGTTA CTATGGCGCT CTGGGTGCGA
.....
3151 CACCGGCTCC AGATTTATCA GCAATAAACC AGCCAGCCGG AAGGGCCGAG
GTGCGCGAGG TCTAAATAGT CGTTATTTGG TCGGTGCGCC TTCCCGGCTC
.....
3201 CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT CTATTAATTG
GCGTCTTAC CAGGACGTTG AAATAGGCGG AGGTAGGTCA GATAATTAAC
.....
3251 TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT TTGCGCAACG
AACGGCCCTT CGATCTCAT CATCAAGCGG TCAATTATCA AACGCGTTGC
.....
3301 TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC GTTGGTATG
AACAAACGTA ACGATGTCCG TAGCACCACA GTGCGAGCAG CAAACCATAC
.....
3351 GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA CATGATCCCC
CGAAGTAAGT CGAGGCCAAG GGTGCTAGT TCCGCTCAAT GTACTAGGGG
.....
3401 CATGTTGTGC AAAAAAGCGG TTAGCTCCTT CGGTCCCTCG ATCGTTGTCA
GTACAACAG TTTTTCGCC AATCGAGGAA GCCAGGAGGC TAGCAACAGT
.....
3451 GAAGTAAGTT GGCCGAGTG TTATCACTCA TGGTTATGGC AGCACTGCAT
CTTCATTCAA CCGGCTCAC AATAGTGAGT ACCAATACCG TCGTGACGTA
.....

3501 AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG TGA CTGGTGA
TTAAGAGAAT GACAGTACGG TAGGCATTCT ACGAAAAGAC ACTGACCACT

3551 GTACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCGA CCGAGTTGCT
CATGAGTTGG TTCAGTAAGA CTCTTATCAC ATACGCCGCT GGCTCAACGA

3601 CTGCCCCGGC GTCAATACGG GATAATACCG CGCCACATAG CAGAACTTTA
GAACGGGCGC CAGTTATGCC CTATTATGGC GCGGTGTATC GTCTTGAAT

3651 AAAGTGCTCA TCATTGGAAA ACGTTCTTCG GGGCGAAAAC TCCTCAAGGAT
TTTACAGAGT AGTAACCTTT TGCAAGAAGC CCCGCTTTTG AGAGTTCTTA

ApaLI

3701 CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT GCACCCAACCT
GAATGGCGAC AACTCTAGGT CAAGCTACAT TGGGTGAGCA CGTGGGTGTA

3751 GATCTTCAGC ATCTTTTACT TTCACCAGCG TTTCTGGGTG AGCAAAAACA
CTAGAAGTCG TAGAAAATGA AAGTGGTCCG AAAGACCCAC TCGTTTTGT

3801 GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC GGAATGTGT
CCTTCCGTTT TACGGCGTTT TTTCCCTTAT TCCCGCTGTG CCTTTACAAC

3851 AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT TATCAGGGTT
TTATGAGTAT GAGAAGGAAA AAGTTATAAT AACTTCGTAA ATAGTCCCAA

3901 ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAACAA
TAACAGAGTA CTCGCCTATG TATAAATTA CATAAATCTT TTTATTGT

3951 ATAGGGGTTT CGCGCACATT TCCCCGAAA GTGCCACCTG ACGTCTAAGA
TATCCCCAAG GCGCGTGTA AGGGGCTTTT CACGGTGGAC TGCAGATTCT

4001 AACCATTATT ATCATGACAT TAACCTATAA AAATAGGCGT ATCAGGAGGC
TTGGTAATAA TAGTACTGTA ATTGGATATT TTTATCCGCA TAGTGCTCCG

4051 CCTTTCGTCT CGCGCGTTTC GGTGATGACG GTGAAAACCT CTGACACATG
GGAAGCAGA GCGCGCAAAG CCACTACTGC CACTTTTGA GACTGTGTAC

4101 CAGCTCCCGG AGACGGTCAC AGCTTGTCTG TAAGCGGATG CCGGGAGCAG
GTCGAGGGCC TCTGCCAGTG TCGAACAGAC ATTGCCTTAC GGCCCTCGTC

4151 ACAAGCCCGT CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT CGGGGCTGGC
TGTTCCGGCA GTCCCGCGCA GTCGCCACA ACCGCCACA GCCCGACCG

ApaLI

4201 TTAATATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA CCATATCGAC
AATTGATACG CCGTAGTCTC GTCTAACATG ACTCTCACGT GGTATAGCTG

4251 GCTCTCCCTT ATGCGACTCC TGCATTAGGA AGCAGCCCAG TAGTAGGTTG
CGAGAGGGAA TACGCTGAGG ACGTAATCCT TCGTCGGGTC ATCATCCAAC

4301 AGGCCGTTGA GCACCGCCGC CGCAAGGAAT GGTGCATGCA AGGAGATGGC
TCCGGCAACT CGTGCGGCG GCGTTCTTA CCACGTACGT TCCTCTACCG

4351 GCCCAACAGT CCCCCGGCCA CGGGGCCTGC CACCATACCC ACGCCGAAC
CGGGTTGTCA GGGGCGCGT GCCCGGACG GTGGTATGGG TGCGGCTTGT

4401 AAGCACTAAT AGGAATTGAT TTGGATGGTA TAAACGAAA CAAAAAAG
TTCGTGATTA TCCTTAAC TAACCTACCAT ATTTGCCTTT GTTTTTTTC

4451 AGCTGGTACT ACTTCTTTT AAATTATTTT ATTATTTGAT TTTATTTAAT
TCGACCATGA TGAAAGAAAT TTTAATAAAA TAATAAACTA AAATAAAATTA

4501 AGTATATATT ATATTTTGAA CGTAGATTAT TTTGTTGAAA GTTGCTGTAG
TCATATATAA TATAAAACTT GCATCTAATA AAACAACCTT CAACGACATC

4551 TGCCATTGAT TCGTAACACT AATTCTGTAT TAGTCATTCC TCTTGTTTGA
ACGGTAACTA AGCATTGTGA TTAAGACATA ATCAGTAAGG AGAACAACT

4601 TAGTATCCAA AAAACGGCT ATTTTGTTC AATCTTATTT CCTGCATATT
ATCATAGGTT TTTTGGCGA TAAAAAACG TTAGAATAAA GGACGTATAA

4651 ATACAGATAA CATAATGAAA GAAAAATCT TTTTTTTGT TCTTCAATGA
TATGTCTATT GTATTACTTT CTTTTTTGA AAAAAACA AGAAGTTACT

4701 TGATTTCAAC CATTCTTTTA AACATTGATC AATCTCTGAG CAACAACCCC
ACTAAAGTTG GTAAGAAAT TTGTAAC TAGTAAAGGACTC GTTGTGGGG

4751 ATACACACTG GTTTATATAC CGCCCTTTT ACAGTTGAAG AAAGAAATAG
TATGTGTGAC CAAATATATG CGGGGAAAA TGTCACCTC TTTCTTTATC

4801 AAATAGAAAT AGCAAACAAA AGATATGACA GTCAACACTA AGACCTATAG
TTTATCTTTA TCGTTTGTTC TCTATACTGT CAGTTGTGAT TCTGGATATC

4851 TGAGAGAGCA GAAACTCATG CCTCACCAGT AGCACAGCGA TTATTTGAT
ACTCTCTCGT CTTTGAGTAC GGAGTGGTCA TCGTGTGCT AATAAAGCTA

4901 TAATGGAAC GAAGAAAACC AATTATGTG CATCAATTGA CGTTGATACC
ATTACCTTGA CTCTTTTGG TTAATACAC GTAGTTAACT GCAACTATGG

AvaI

4951 ACTAAGGAGT TCCTCGAGTT AATTGATAAA TTAGGTCCTT ATGTATGCTT
TGATTCTCA AGGAGCTCAA TTAATATTT AATCCAGGAA TACATACGAA

5001 AATCAAGACT CATATTGATA TAATCAATGA TTTTCTCTAT GAATCCACTA
TTAGTTCTGA GTATAACTAT ATTAGTTACT AAAAAGGATA CTTAGGTGAT

5051 TTGAACCATT ATTAGAACTT TCACGTAAAC ATCAATTTAT GATTTTGA
AACTTGGTAA TAATCTTGAA AGTGCATTG TAGTTAATA CTAAAACTT

5101 GATAGAAAAT TTGCTGATAT TGGTAATACC GTAAAGAAAC AATATATTGG
CTATCTTTTA AACGACTATA ACCATTATGG CATTTCTTTG TTATATAACC

5151 TGGAGTTTAT AAAATTAGTA GTTGGGCAGA TATTACCAAT GCTCATGGTG
ACCTCAAATA TTTAATCAT CAACCCGTCT ATAATGGTTA CGAGTACCAC

5201 TCACTGGGAA TGGAGTGGTT GAAGGATTAA AACAGGGAGC TAAAGAAACC
AGTGACCCCT ACCTACCAA CTTCCTAATT TTGTCCTCG ATTTCTTTGG

5251 ACCACCAACC AAGAGCCAAG AGGGTTATTG ATGTTAGCTG AATTATCATC
TGGTGGTTGG TTCTCGGTC TCCAATAAC TACAATCGAC TTAATAGTAG

5301 AGTGGGATCA TTAGCATATG GAGAATATTC TCAAAAACT GTTGAAATTG
TCACCCTAGT AATCGTATAC CTCTTATAAG AGTTTTTTGA CAACTTTAAC

5351 CTAATCCGA TAAGGAATTT GTTATTGGAT TTATTGCCA ACGTGATATG
GATTTAGGCT ATTCCTTAAA CAATAACCTA AATAACGGGT TGCATATAC

5401 GGTGGCCAAG AAGAAGGATT TGATTGGCTT ATTATGACAC CTGGAGTTGG
CCACCGGTTT TTCTTCCTAA ACTAACCGAA TAATACTGTG GACCTCAACC
.....
5451 ATTAGATGAT AAAGGTGATG GATTAGGACA ACAATATAGA ACTGTTGATG
TAATCTACTA TTCCACTAC CTAATCCTGT TGTATATCT TGACAACTAC
.....
5501 AAGTTGTTAG CACTGGAACT GATATTATCA TTGTTGGTAG AGGATTGTTT
TTCAACAATC GTGACCTTGA CTATAATAGT AACAACCATC TCCTAACAAA
.....
5551 GGTAAAGGAA GAGATCCAGA TATTGAAGGT AAAAGGTATA GAAATGCTGG
CCATTTCTCT CTCTAGGTCT ATAACCTCCA TTTTCCATAT CTTTACGACC
.....
5601 TTGGAATGCT TATTTGAAAA AGACTGGCCA ATTATAAATG TGAAGGGGGA
AACCTTACGA ATAAACTTTT TCTGACCGGT TAATATTAC ACTTCCCCCT
.....
5651 GATTTTCACT TTATTAGATT TGTATATATG TAGAATAAAT AAATAAATAA
CTAAAAGTGA AATAATCTAA ACATATATAC ATCTTATTTA TTTATTTATT
.....
5701 GTTAAATAAA TAATTAAATA AGGGTGGTAA TTATTACTAT TTACAATCAA
CAATTTATTT ATTAATTTAT TCCCACCATT AATAATGATA AATGTTAGTT
.....
5751 AGGTGGTCCT TCTAGCTGTA ATCCGGGCAG CGCAACGGAA CATTTCATCAG
TCCACCAGGA AGATCGACAT TAGGCCCGTC GCGTTGCCCT GTAAGTAGTC
.....
5801 TGTAAAAATG GAATCAATAA AGCCCTGCGC TCATGAGCCC GAAGTGGCGA
ACATTTTAC CTTAGTTATT TCGGGACGCG AGTACTCGGG CTTCACCGCT
.....
5851 GCCCGATCTT CCCCATCGGT GATGTCGGCG ATATAGGCGC CAGCAACCGC
CGGGCTAGAA GGGGTAGCCA CTACAGCCGC TATATCCGCG GTCGTTGGCG
.....
5901 ACCTGTGGCG CCGCAGCGCG CAGGGTCAGC CTGAATACGC GTTTAATGAC
TGGACACCGC GCGTCGCGC GTCCCAGTCG GACTTATGCG CAAATTACTG
.....
5951 CAGCACAGTC GTGATGGCAA GGTGAGAATA GCCCAAGTCG GCCGAGGGGC
GTCGTGTCAG CACTACCGTT CCAGTCTTAT CGGGTTCAGC CGGCTCCCCG
.....
6001 CTGTACAGTG AGGGAAGATC TGATATTGAC GAAGAGGAAC CAATGTAACG
GACATGTCAC TCCCTTCTAG ACTATAACTG CTCTCCTTG GTTACATTGC
.....
6051 TTACACTGAA GAAAACACAC AATAAACGGG AAGAAACGGT GTAAAAGTGT
AATGTGACTT CTTTGTGTG TTATTGCCC TTCTTTGCCA CATTTCACA
.....
6101 GAAAATAATT TTTGAATATC ATTTCCCTTG GTTTAATTCC AAACGAAACG
CTTTTATTAA AAACCTATAG TAAAGGGAAC CAAATTAAGG TTTGCTTTGC
.....

EcoRI

6151 TGTTTTTTTT AGAGAATGGG AATTCTTATT GGATGTCTAG ATTGTTTGTT
ACAAAAAAA TCTCTTACCC TTAAGAATAA CCTACAGATC TAACAAACAA
.....

ApaLI

6201 TACTCCAGAC TGTGCACAAA AACGTTTGA TGGATGATCA GAAGATATTT
ATGAGGTCTG ACACGTGTTT TTGCAACCT ACCTACTAGT CTCTATAAA
.....
6251 TTAGGCTTAG CTCTAAATAT AAGAAATGAT GCTTGAAAAA CCAGACAGAA
AATCCGAATC GAGATTTATA TTCTTACTA CGAATTTTTT GGTCTGTCTT
.....
6301 ATTGAGTTTC AAAAATTGGT AATGTGAGGT ATTAGTCAAC TAACCAAATA
TAACTCAAAG TTTTAAACCA TTACACTCCA TAATCAGTTG ATTGGTTTAT
.....

6351 ACAATGCAAA CCGGTGATA CATTTCATT TGAAAATAAT GAAACTGGAA
TGTTACGTTT GGCCAACTAT GTAAAGTAAA ACTTTTATTA CTTTGACCTT

6401 TTGGATGACC AGCACACAAA CACATAAAGT AATTATGGGA ATTAGAAGCG
AACCTACTGG TCGTGTGTTT GTGTATTTC TTAATACCCT TAATCTTCGC

6451 AACATAGAGG AGTACTTGGC CACGAACAGA ATACAAGTGG GAACACTATT
TTGTATCTCC TCATGAACCG GTGCTTGTCT TATGTTTACC CTTGTGATAA

6501 TTCTCCATTG TTTTAGTTCT GTTTTTTGT CAGCCTAGTT TTGTGCTATG
AAGAGGTAAC AAAATCAAGA CAAAAAACA GTCGGATCAA AACACGATAC

HindIII

6551 TGTAAGAAAT ATTGCCAAGA AAAAAAGCTT GTTTTGTGGC CAGTGTCCGA
ACATTTTTTA TAACGGTTCT TTTTTCGAA CAAAACACCG GTCACAGGCT

6601 AAAAAATTTT GGGGAATCTT CGGATTAATT TATGTTTCA
TTTTTTAAAA CCCCTTAGAA GCCTAATTAA ATACAAAAGT

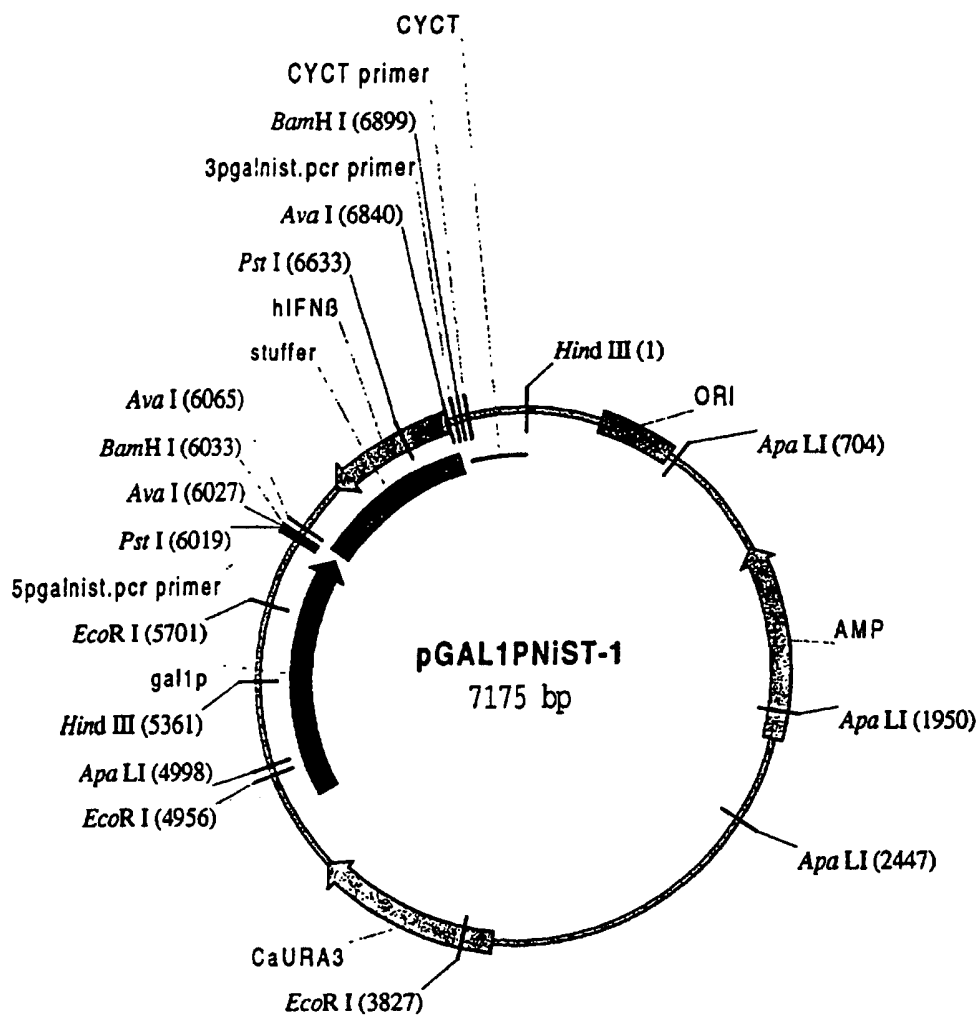


Fig 64

HindIII

1 AGCTTGAGTA TTCTATAGTG TCACCTAAAT AGCTTGGCGT AATCATGGTC
TCGAACATCAT AAGATATCAC AGTGGATTTA TCGAACCACA TTAGTACCAG
.....
51 ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT CCACACAACA
TATCGACAAA GGACACACTT TAACAATAGG CGAGTGTAA GGTGTGTTGT
.....
101 TACGAGCCGG AAGCATAAAG TGTAAAGCCT GGGGTGCCTA ATGAGTGAGC
ATGCTCGGCC TCGTATTTC ACATTTCGGA CCCACGGAT TACTCACTCG
.....
151 TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA
ATTGAGTGTA ATTAACGCAA CGCGAGTGAC GGGCGAAAGG TCAGCCCTTT
.....
201 CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG
GGACAGCAGC GTCGACGTAA TTACTTAGCC GGTTCGCGCG CCTCTCCGC
.....
251 GTTTGCGTAT TGGGCGCTCT TCCGCTTCCT CGCTCACTGA CTCGCTGCGC
CAAACGCATA ACCCGCGAGA AGCGAAGGA GCGAGTGA CTGAGCAGCG
.....
301 TCGGTCGTTT GGCTGCGGCG AGCGGTATCA GCTCACTCAA AGGCGGTAAT
AGCCAGCAAG CCGACGCCGC TCGCCATAGT CGAGTGAGTT TCCGCCATTA
.....
351 ACGGTTATCC ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA
TGCCAATAGG TGTCTTAGTC CCCTATTGCG TCCTTTCTTG TACACTCGTT
.....
401 AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT
TTCCGGTCGT TTCCGGTCC TTGGCATTTT TCCGGCGCAA CGACCGCAA
.....
451 TTCCATAGGC TCCGCCCCC TGACGAGCAT CACAAAAATC GACGCTCAAG
AAGGTATCCG AGGCGGGGG ACTGCTCGTA GTGTTTCTAG CTGCGAGTTC
.....
501 TCAGAGGTGG CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC
AGTCTCCACC GCTTTGGGCT GTCCTGATAT TTCTATGGTC CGCAAAGGGG
.....
551 CTGGAAGCTC CCTCGTGCGC TCTCCTGTTT CGACCCTGCC GCTTACCGGA
GACCTTCGAG GGAGCAGCG AGAGGACAAG GCTGGGACGG CGAATGGCCT
.....
601 TACCTGTCCG CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCATAGCTC
ATGGACAGGC GGAAAGAGGG AAGCCCTTCG CACCGCGAAA GAGTATCGAG
.....
651 ACGCTGTAGG TATCTCAGTT CCGTGTAGGT CGTTCGCTCC AAGCTGGGCT
TGCGACATCC ATAGAGTCAA GCCACATCCA GCAAGCGAGG TTCGACCGA
.....

ApaLI

701 GTGTGCACGA ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC
CACACGTGCT TGGGGGGCAA GTCGGGCTGG CGACGCGGAA TAGGCCATTG
.....
751 TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC
ATAGCAGAAC TCAGGTTGGG CCATTCTGTG CTGAATAGCG GTGACCGTCG
.....
801 AGCCACTGGT AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG
TCGGTGACCA TTGTCTAAT CGTCTCGCTC CATACTCCG CCACGATGTC
.....
851 AGTTCTTGAA GTGGTGGCCT AACTACGGCT AACTAGAAG GACAGTATTT
TCAAGAACTT CACCACCGGA TTGATGCCGA TGTGATCTC CTGTCATAAA
.....
901 GGTATCTGCG CTCTGCTGAA GCCAGTTACC TTCGGAAAAA GAGTTGGTAG
CCATAGACGC GAGACGACTT CGGTCAATGG AAGCCTTTTT CTCAACCATC
.....

Fig 65

951 CTCTTGATCC GGCAAACAAA CCACCGCTGG TAGCGGTGGT TTTTGTGTT
GAGAACTAGG CCGTTTGTGTT GGTGGCGACC ATCGCCACCA AAAAAACAAA
.....
1001 GCAAGCAGCA GATTACGCGC AGAAAAAAG GATCTCAAGA AGATCCTTG
CGTTCGTCGT CTAATGCGCG TCTTTTTC CTAGAGTTCT TCTAGGAAAC
.....
1051 ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAACT CACGTTAAGG
TAGAAAAGAT GCCCCAGACT GCGAGTCACC TTGCTTTTGA GTGCAATTCC
.....
1101 GATTTTGGTC ATGAGATTAT CAAAAGGAT CTTACCTAG ATCCTTTTAA
CTAAAACCAG TACTCTAATA GTTTTCTTA GAAGTGGATC TAGGAAAATT
.....
1151 ATTAATAATG AAGTTTAAA TCAATCTAAA GTATATATGA GTAACTTGG
TAATTTTAC TTCAAAATT AGTTAGATT CATATATACT CATTTGAACC
.....
1201 TCTGACAGTT ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG
AGACTGTCAA TGTTACGAA TTAGTCACTC CGTGGATAGA GTCGCTAGAC
.....
1251 TCTATTTCGT TCATCCATAG TTGCCTGACT CCGCTCGTG TAGATAACTA
AGATAAAGCA AGTAGGTATC AACGACTGA GGGGACGAC ATCTATTGAT
.....
1301 CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGGA
GCTATGCCCT CCCGAATGGT AGACCGGGT CACGACGTTA CTATGGCGCT
.....
1351 GACCCACGCT CACCGGCTCC AGATTTATCA GCAATAAACC AGCCAGCCGG
CTGGGTGCGA GTGGCCGAGG TCTAAATAGT CGTTATTGG TCGTCCGCC
.....
1401 AAGGGCCGAG CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT
TTCCCGCTC GCGTCTTAC CAGGACGTTG AAATAGGCGG AGGTAGGTCA
.....
1451 CTATTAATTG TTGCCGGGA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT
GATAATTAAC AACGGCCCTT CGATCTCATT CATCAAGCGG TCAATTATCA
.....
1501 TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC
AACGCGTTGC AACACCGTA ACGATGTCCG TAGCACCACA GTGCGAGCAG
.....
1551 GTTTGGTATG GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA
CAAACCATAC CGAAGTAAGT CGAGGCCAAG GGTGCTAGT TCCGCTCAAT
.....
1601 CATGATCCCC CATGTTGTGC AAAAAAGCGG TTAGCTCCTT CGGTCTCCG
GTACTAGGGG GTACAACACG TTTTTCGCC AATCGAGGAA GCCAGGAGC
.....
1651 ATCGTTGTCA GAAGTAAGTT GGCCGAGTG TTATCACTCA TGGTTATGGC
TAGCAACAGT CTTCAATCAA CCGCGTCAC AATAGTAGT ACCAATACG
.....
1701 AGCACTGCAT AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG
TCGTGACGTA TTAAGAGAAT GACAGTACGG TAGGCATTCT ACGAAAAGAC
.....
1751 TGACTGGTGA GACTCAACC AAGTCATTCT GAGAATAGTG TATCGGCGA
ACTGACCACT CATGAGTTGG TTCAGTAAGA CTCTTATCAC ATACGCCGCT
.....
1801 CCGAGTTGCT CTTGCCCGG GTCAATACGG GATAATACCG CGCCACATAG
GGCTCAACGA GAACGGGCG CAGTTATGCC CTATTATGGC GCGGTGTATC
.....
1851 CAGAACTTTA AAAGTGCTCA TCATTGAAA ACGTTCTTCG GGGCGAAAAC
GTCTTGAAAT TTACAGAGT AGTAACCTTT TGCAAGAAGC CCGCTTTTG
.....

ApaLI

1901 TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT
AGAGTTCCTA GAATGGCGAC AACTCTAGGT CAAGCTACAT TGGGTGAGCA

ApaLI

1951 GCACCCAACT GATCTTCAGC ATCTTTTACT TTCACCAGCG TTTCTGGGTG
CGTGGGTTGA CTAGAAGTCG TAGAAAATGA AAGTGGTCCG AAAGACCCAC

2001 AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGAATA AGGCGGACAC
TCGTTTTTGT CCTTCCGTTT TACGGCGTTT TTTCCCTTAT TCCCGCTGTG

2051 GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT
CCTTTACAAC TTATGAGTAT GAGAAGGAAA AAGTTATAAT AACTTCGTAA

2101 TATCAGGGTT ATTGTCTCAT GAGCGGATAC ATATTGGAAT GTATTTAGAA
ATAGTCCCAA TAACAGAGTA CTCGCCTATG TATAAACTTA CATAAATCTT

2151 AAATAAACAA ATAGGGGTTT CGCGCACATT TCCCCGAAAA GTGCCACCTG
TTTATTGTGT TATCCCCAAG GCGCGTGTA AGGGGCTTTT CACGGTGGAC

2201 ACGTCTAAGA AACCATTATT ATCATGACAT TAACCTATAA AAATAGGCGT
TGCAGATTCT TTGGTAATAA TAGTACTGTA ATTGGATATT TTTATCCGCA

2251 ATCAGGAGGC CCTTTCGTCT CGCGCGTTTC GGTGATGACG GTGAAACCT
TAGTGCTCCG GGAAGCAGA GCGCGCAAAG CCACTACTGC CACTTTTGGG

2301 CTGACACATG CAGCTCCCGG AGACGGTCAC AGCTTGTCTG TAAGCGGATG
GACTGTGTAC GTCGAGGGCC TCTGCCAGTG TCGAACAGAC ATTCCGCTAC

2351 CCGGGAGCAG ACAAGCCCGT CAGGGCGCGT CAGCGGGTGT TGGCGGGTGT
GGCCCTCGTC TGTTCGGGCA GTCCCGCGCA GTCGCCACA ACCGCCACA

ApaLI

2401 CGGGGCTGGC TTAACATATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA
GCCCCGACCG AATTGATACG CCGTAGTCTC GTCTAACATG ACTCTCACGT

ApaLI

2451 CCATATGCGG TGTGAAATAC CGCACAGATG CGTAAGGAGA AAATACCGCA
GGTATACGCC ACACCTTTATG GCGTGTCTAC GCATTCTCTT TTTATGGCGT

2501 TCAGGCGAAA TTGTAAACGT TAATATTTTG TTAATTCG CGTTAAATAT
AGTCCGCTTT AACATTGCA ATTATAAAAC AATTTTAAGC GCAATTATATA

2551 TTGTAAATC AGCTCATTTT TTAACCAATA GGCCGAAATC GGCAAAATCC
AACAATTTAG TCGAGTAAAA AATTGGTTAT CCGGCTTTAG CCGTTTATAGG

2601 CTTATAATC AAAAGAATAG ACCGAGATAG GGTGAGTGT TGTTCAGTT
GAATATTTAG TTTTCTTATC TGGCTCTATC CCACTCACA ACAAGGTCAA

2651 TGGAACAAGA GTCCACTATT AAAGAACGTG GACTCCAACG TCAAAGGGCG
ACCTTGTCT CAGGTGATAA TTTCTTGCAC CTGAGGTTGC AGTTTCCCGC

2701 AAAAACCGTC TATCAGGGCG ATGGCCCACT ACGTGAACCA TCACCCAAAT
TTTTTGGCAG ATAGTCCCGC TACCGGGTGA TGCATTGGT AGTGGGTTTA

2751 CAAGTTTTTT GCGGTCGAGG TGCCGTAAG CTCTAAATCG GAACCCTAAA
GTTCAAAAAA CGCCAGCTCC ACGGCATTTC GAGATTTAGC CTGGGATTT

2801 GGGAGCCCC GATTTAGAGC TTGACGGGA AAGCCGGCGA ACGTGGCGAG
CCCTCGGGG CTAATCTCG AACTGCCCT TTCGGCCGCT TGCACCGCTC
.....
2851 AAAGGAAGG AAGAAAGCGA AAGGAGCGG CGCTAGGGCG CTGGCAAGTG
TTTCTTCCC TTCCTTCGCT TTCCTCGCC GCGATCCCGC GACCGTTCAC
.....
2901 TAGCGGTCAC GCTGCGCGTA ACCACCACAC CCGCCGCGCT TAATGCGCCG
ATCGCCAGTG CGACGCGCAT TGGTGGTGTG GCGGCGCGA ATTACGCGGC
.....
2951 CTACAGGGCG CGTCCATTCTG CCATTGAGC TCGCAACTG TTGGGAAGGG
GATGTCCCGC GCAGGTAAGC GGTAAGTCCG ACGCGTTGAC AACCTTCCC
.....
3001 CGATCGGTGC GGGCCTCTTC GCTATTACGC CAGCTGGCGA AAGGGGGATG
GCTAGCCACG CCCGAGAAG CGATAATGCG GTCGACCGCT TTCCCTCTAC
.....
3051 TGCTGCAAGG CGATTAAGTT GGGTAACGCC AGGGTTTTC CAGTCACGAC
ACGACGTTC GCTAATTCAA CCCATTGCGG TCCCAAAAGG GTCAGTGCTG
.....
3101 GTTGTAAGAC GACGGCCAGT GAATGTAAAT ACGACTCACT ATAGGGCGAA
CAACATTTTG CTGCCGGTCA CTTAACATTA TGCTGAGTGA TATCCCGCTT
.....
3151 TTGGTTTTC AATGATGAGC ACTTTTAAAG TTCGTCTATG TGGCGGGTA
AACCAAAAGG TTAATACTCG TGAAAATTC AAGACGATAC ACCGCGCCAT
.....
3201 TTATCCGTG TTGACGCGG GCAAGAGCAA CTCGGTCGCC GCATACACTA
AATAGGGCAC AACTGCGGCC CGTCTCGTT GAGCCAGCGG CGTATGTGAT
.....
3251 TTCTCAGAAT GACTGGTTG AGTACTAATA GGAATTGATT TGGATGGTAT
AAGAGTCTTA CTGAACCAAC TCATGATTAT CCTTAATAA ACCTACCATA
.....
3301 AAACGGAAC AAAAAAGA GCTGGTACTA CTTCTTTTAA AATTATTTTA
TTTGCTTTG TTTTCTTC CGACCATGAT GAAAGAAATT TTAATAAAAT
.....
3351 TTATTTGATT TTATTTAATA GTATATATTA TATTTTGAAC GTAGATTATT
AATAAACTAA AATAAATTAT CATATATAAT ATAAACTTG CATCTAATAA
.....
3401 TTGTTGAAAG TTGCTGTAGT GCCATTGATT CGTAACACTA ATTCTGTATT
AACAACCTTC AACGACATCA CGGTAATAA GCATTGTGAT TAAGACATAA
.....
3451 AGTCATTCTT CTGTTTGAT AGTATCCAA AAAACGGCTA TTTTGTGCA
TCAGTAAGGA GAACAACTA TCATAGTTT TTTGCGGAT AAAAAACGT
.....
3501 ATCTTATTTT CTGCATATTA TACAGATAAC ATAATGAAAG AAAAAATCTT
TAGAATAAAG GACGTATAAT ATGTCTATTG TATTACTTTC TTTTGTAGAA
.....
3551 TTTTGTGTT CTCAATGAT GATTTCACAC ATTCTTTTAA ACATTGATCA
AAAAAACAA GAAGTTACTA CTAAAGTTGG TAAGAAATTT GTAACTAGT
.....
3601 ATTCTGAGC AACAACCCCA TACACACTGG TTTATATACC GCCCCTTTTA
TAAGGACTCG TTGTTGGGT ATGTGTGACC AAATATATGG CGGGGAAAT
.....
3651 CAGTTGAAGA AAGAAATAGA AATAGAAATA GCAACAAAA GATATGACAG
GTCAACTTCT TTCTTTATCT TTATCTTTAT CGTTGTTTT CTATACTGTC
.....
3701 TCAACACTAA GACCTATAGT GAGAGAGCAG AAATCATGC CTCACCACTA
AGTTGTGATT CTGGATATCA CTCTCTGTC TTTGAGTACG GAGTGGTCAT
.....
3751 GCACAGCGAT TATTTGATT AATGGAAGT AAGAAAACCA ATTTATGTGC
CGTGTGCTA ATAAAGCTAA TTACCTTGAC TTCTTTTGGT TAAATACAG
.....

EcoRI

3801 ATCAATTGAC GTTGATACCA CTAAGGAATT CCTTGAATTA ATTGATAAAT
TAGTTAACTG CAACTATGGT GATTCTTAA GGAAGCTTAAT TAACTATTTA
.....
3851 TAGGTCCTTA TGTATGCTTA ATCAAGACTC ATATTGATAT AATCAATGAT
ATCCAGGAAT ACATACGAAT TAGTCTGAG TATAACTATA TTAGTTACTA
.....
3901 TTTTCCTATG AATCCACTAT TGAACCATT AATTAGACTT CACGTAAACA
AAAAGGATAC TTAGGTGATA ACTTGGTAAT AATCTTGAAA GTGCATTTGT
.....
3951 TCAATTTATG ATTTTGAAG ATAGAAAATT TGCTGATATT GGTAATACCG
AGTTAAATAC TAAAACTTC TATCTTTTAA ACGACTATAA CCATTATGGC
.....
4001 TAAAGAAACA ATATATTGGT GGAGTTTATA AAATTAGTAG TTGGGCAGAT
ATTTCCTTGT TATATAACCA CCTCAAATAT TTAATCATC AACCCGTCTA
.....
4051 ATTACCAATG CTCATGGTGT CACTGGGAAT GGAGTGGTTG AAGGATTAA
TAATGGTTAC GAGTACCACA GTGACCCTTA CCTCACCAC TTCCTAATTT
.....
4101 ACAGGGAGCT AAAGAAACCA CCACCAACCA AGAGCCAAGA GGGTTATTGA
TGTCCCTCGA TTTCTTTGGT GGTGGTTGGT TCTCGGTCT CCCAATAACT
.....
4151 TGTTAGCTGA ATTATCATCA GTGGGATCAT TAGCATATGG AGAATATTCT
ACAATCGACT TAATAGTAGT CACCCTAGTA ATCGTATACC TCTTATAAGA
.....
4201 CAAAAAAGT TGAATTTGC TAAATCCGAT AAGGAATTTG TTATTGGATT
GTTTTTTGAC AACTTTAAGC ATTTAGGCTA TCCCTTAAAC AATAACCTAA
.....
4251 TATTGCCCAA CGTGATATGG GTGGCCAAGA AGAAGGATTT GATTGGCTTA
ATAACGGGTT GCACTATACC CACCGGTCT TCTTCCTAAA CTAACCGAAT
.....
4301 TTATGACACC TGGAGTTGGA TTAGATGATA AAGGTGATGG ATTAGGACAA
AATACTGTGG ACCTCAACCT AATCTACTAT TTCCACTACC TAATCCTGTT
.....
4351 CAATATAGAA CTGTTGATGA AGTTGTTAGC ACTGGAAGT ATATTATCAT
GTTATATCTT GACAACTACT TCAACAATCG TGACCTTGAC TATAATAGTA
.....
4401 TGTTGGTAGA GGATTGTTTG GTAAAGGAAG AGATCCAGAT ATTGAAGGTA
ACAACCATCT CCTAACAAAC CATTTCTTC TCTAGTCTA TAACTTCCAT
.....
4451 AAAGGTATAG AAATGCTGGT TGAATGCTT ATTTGAAAAA GACTGGCCAA
TTTCCATATC TTTACGACCA ACCTTACGAA TAACTTTTT CTGACCGGTT
.....
4501 TTATAAATGT GAAGGGGAG ATTTTCACTT TATTAGATTT GTATATATGT
AATATTTACA CTTCCTCTC TAAAAGTGAA ATAATCTAAA CATATATACA
.....
4551 AGAATAAATA AATAAATAAG TTAATAAAT AATTAAATAA GGGTGGTAAT
TCTTATTAT TATTTATTC AATTTATTTA TTAATTTATT CCCACATTA
.....
4601 TATTACTATT TACAATCAAA GGTGGTCCTT CTAGCTGTAA TCCGGGCAGC
ATAATGATAA ATGTTAGTTT CCACCAGGAA GATCGACATT AGGCCGTCG
.....
4651 GCAACGGAAC ATTCATCAGT GTAAAAATGG AATCAATAAA GCCCTGCGCA
CGTTGCCCTG TAAGTAGTCA CATTTTACC TTAGTTATTT CGGGACGCGT
.....
4701 GCGCGCAGGG TCAGCCTGAA TACGCGTTTA ATGACCAGCA CAGTCGTGAT
CGCGCGTCCC AGTCGGACTT ATGCGCAAAT TACTGGTCGT GTCAGCACTA
.....

4751 GGCAAGGTCA GAATAGCCCA AGTCGGCCGA GGGGCTGTGA CAGTGAGGGA
CCGTTCCAGT CTATCGGGT TCAGCCGGCT CCCCAGACAT GTCACTCCCT

4801 AGATCTGATA TTGACGAAGA GGAACCAATG TAACGTTACA CTGAAGAAAA
TCTAGACTAT AACTGCTTCT CCTTGGTTAC ATTGCAATGT GACTTCTTTT

4851 CACACAATAA ACGGGAAGAA ACGGTGTAAG AGTGTGAAAA TAATTTTGA
GTGTGTTATT TGCCCTTCTT TGCCACATTT TCACACTTTT ATTAAAACT

4901 ATATCATTTT CCTTGGTTTA ATTCCAAACG AAACGTGTTT TTTTATGAGA
TATAGTAAAG GGAACCAAAT TAAGGTTTGC TTTGCACAAA AAAAATCTCT

EcoRI

ApaLI

4951 ATGGGAATTC TTATTGGATG TCTAGATTGT TTGTTTACTC CAGACTGTGC
TACCCTTAAG AATAACCTAC AGATCTAACA AACAAATGAG GTCTGACACG

ApaLI

5001 ACAAAAACGT TTGGATGGAT GATCAGAAGA TATTTTATAG CTTAGCTCTA
TGTTTTTGCA AACCTACCTA CTAGTCTTCT ATAAAAATCC GAATCGAGAT

5051 AATATAAGAA ATGATGCTTG AAAAACCAGA CAGAAATTGA GTTTCAAAAA
TTATATTCTT TACTACGAAC TTTTGGTCT GTCTTTAACT CAAAGTTTTT

5101 TTGGTAATGT GAGGTATTAG TCAACTAACC AAATAACAAT GCAAACCGGT
AACCATTACA CTCATAATC AGTTGATTGG TTTATTGTTA CGTTGGCCA

5151 TGATACATTT CATTTTGAAA ATAATGAAAC TGGAATTGGA TGACCAGCAC
ACTATGTAAA GTAAACTTTT TATTACTTTG ACCTTAACCT ACTGGTCGTG

5201 ACAAACACAT AAAGTAATTA TGGGAATTAG AAGCGAACAT AGAGGAGTAC
TGTTTGTTGA TTTCATTAAT ACCCTTAATC TTCGCTTGTA TCTCTCATG

5251 TTGGCCACGA ACAGAATACA AGTGGGAACA CTATTTTCTC CATTTGTTTTA
AACCGGTGCT TGTCTTATGT TCACCCTTGT GATAAAGAG GTAACAAAAT

5301 GTTCTGTTTT TTTGTCAGCC TAGTTTTGTG CTATGTGTAA AAAATATTGC
CAAGACAAAA AAACAGTCGG ATCAAAACAC GATACACATT TTTTATAACG

HindIII

5351 CAAGAAAAAA AGCTTGTTTT GTGGCCAGTG TCCGAAAAAA ATTTTGGGGA
GTTCTTTTTT TCGAACAAAA CACCGGTCAC AGGCTTTTTT TAAAACCCCT

5401 ATCTTCGGAT TAATTTATGT TTTCATTCCA TCGGGGAAAG TGGGGGGGAA
TAGAAGCCTA ATTAAATACA AAAGTAAGGT AGCCCTTTTC ACCCCCCCTT

5451 AAAATTTTAA GCAGTTCACA AAACCTTCCA AAAATATAT GGACAAAGAT
TTTTAAATTT CGTCAAGTGT TTTGGAAGGT TTTTATATA CCTGTTTCTA

5501 GATTGTATTT TCCGACACC AAAATCATAA TTAATTATGA GAAAGTTAAA
CTAACATAAA AGGGCTGTGG TTTTAGTATT AATTAATACT CTTCAATTT

5551 TGTAACGTTA CAATTTATGT TTATTTGAAG GTGAAAAGCG ATTTATGATT
ACATTGCAAT GTTAAATACA AATAAACTTC CACTTTTCGC TAAATACTAA

5601 TTTCCGAAAT GAAAATTTTT TTTAGGTTTA TTTTMTTGT CGGGCAAAGA
AAAGGCTTTA CTTTAAAAAA AAATCCAAAT AAAAAAACA GCCCGTTTCT

EcoRI

5651 AAAACTGAAC AAGGATTAT AAAATTTTGT GTGTTTGTGT GTGTCTGGAG
TTTTGACTTG TTCCTAATAA TTTTAAAAAC CACAAACAAA CACAGACCTC

EcoRI

5701 AATTCATTCC TCTCTCATCT TCACACAATG TTTAGACATC TGACACGATT
TTAAGTAAGG AGAGAGTAGA AGTGTGTTAC AAATCTGTAG ACTGTGCTAA

5751 CATGATAGTT CGGTTTCCGG GGTGGTGTGT TAGTTTTCGT TTTCTTTTTT
GTACTATCAA GCCAAAGGCC CCAACCACAA ATCAAAAGCA AAAAGAAAAA

5801 TTTTGAAAG AATGTTTGTAG CTCATTGGTT TTCTTTCTTC ATTCAATAGT
AAAACCTTTC TTACAAAATC GAGTAACCAA AAGAAAGAAG TAAGTTATCA

5851 TTTGAAAGAA TTTGCCCACT TGTATTACA ATCATATAAA ATTAACTTT
AAACTTTCTT AAACGGGTGA ACAATAATGT TAGTATATTT TAATTTGAAA

5901 GATATAAAT AGAGTTTGAA AGTTTCCCAG ATCCTTTTGT ATTTCTTTGT
CTATATTTTA TCTCAAACCT TCAAAGGGTC TAGGAAAAAC TAAAGAAACA

5951 AAATTTTTTT TTCTCCACAC TATACACACA TACAAACCGA TTTTATAAG
TTTAAAAAAA AAGAGGGTGT ATATGTGTGT ATGTTTGGCT AAAAATATTC

PstI

AvaI

BamHI

6001 AAAGAGTTAT ACCCTGCAGC TCGACCTCGA GGGATCCGGG CCCTCTAGAT
TTTCTCAATA TGGGACGTCG AGCTGGAGCT CCCTAGGCCC GGGAGATCTA

AvaI

6051 GCGGCCGCTA GGCCTCGAGG GACTTTTGCA CAAAAAATAA TTTATTTTCC
CGCCGGCGAT CCGGAGCTCC CTGAAAACGT GGTMTTATT AAATAAAAGG

6101 AAAATAAAAT TTAATAAAAT AAAAATAACT CATAATTAA TAAAAATTC
TTTTATTTTA AATTATTTTA TTTTATTGA GTATTAAAT ATTTTAAAG

6151 AAAATCTTCT AGTGTCTTT CATATGCAGT ACATTAGCCA TCAGTCACTT
TTTGAAGA TCACAGGAAA GTATACGTCA TGAATCGGT AGTCAGTGAA

6201 AAACAGCATC TGCTGGTTGA AGAATGCTTG AAGCAATGT CCAGTCCCAG
TTGTCTGTAG ACGACCAACT TCTACGAAC TTCGTAAACA GGTCAGGGTC

6251 AGGCACAGGC TAGGAGATCT TCAGTTTCGG AGGTAACCTG TAAGTCTGTT
TCCGTGTCCG ATCCTCTAGA AGTCAAAGCC TCCATTGGAC ATTCAGACAA

6301 AATGAAGTAA AAGTTCCTTA GGATTTCCAC TCTGACTATG GTCCAGGCAC
TTACTTCATT TTCAAGGAA CCTAAAGGTG AGACTGATAC CAGGTCCGTG

6351 AGTGACTGTA CTCCTTGGCC TTCAGGTAAT GCAGAATCCT CCCATAATAT
TCACTGACAT GAGGAACCGG AAGTCCATTA CGTCTTAGGA GGGTATTATA

6401 CTTTTCAGGT GCAGACTGCT CATGAGTTT CCCCTGGTGA AATCTTCTTT
GAAAAGTCCA CGTCTGACGA GTACTCAAAA GGGGACCACT TTAGAAGAAA

6451 CTCCAGTTTT TCTTCCAGGA CTGCTTTCAG ATGGTTTATC TGATGATAGA
GAGGTCAAAA AGAAGGTCTT GACAGAAGTC TACCAATAG ACTACTATCT

6501 CATTAGCCAG GAGGTTCTCA ACAATAGTCT CATTCCAGCC AGTGCTAGAT
GTAATCGGTC CTCAAGAGT TGTATCAGA GTAAGGTCGG TCACGATCTA

6551 GAATCTTGTC TGAAAATAGC AAAGATGTTT TGGAGCATCT CATAGATGGT
CTTAGAACAG ACTTTTATCG TTTCTACAAG ACCTCGTAGA GTATCTACCA

PstI

6601 CAATGCGGCG TCCTCCTTCT GGAAGTGTG CAGCTGCTTA ATCTCCTCAG
GTTACGCCGC AGGAGGAAGA CCTTGACGAC GTCGACGAAT TAGAGGAGTC

6651 GGATGTCAAA GTTCATCCTG TCCTTGAGGC AGTATTCAAG CCTCCCATTC
CCTACAGTTT CAAGTAGGAC AGGAACTCCG TCATAAGTTC GGAGGGTAAG

6701 AATTGCCACA GGAGCTTCTG AACTGAAAA TTGCTGCTTC TTTGTAGGAA
TTAACGGTGT CCTCGAAGAC TGTGACTTTT AACGACGAAG AAACATCCTT

6751 TCCAAGCAAG TTGTAGCTCA TGGAAAGAGC TGTAGTGGAG AAGCACAACA
AGGTTTCGTT AACATCGAGT ACCTTTCTCG ACATCACCTC TTCGTGTGTG

AvaI

6801 GGAGAGCAAT TTGGAGGAGA CACTTGTGTG TCATGTTTCT CGAGGCCTTT
CCTCTCGTTA AACCTCCTCT GTGAACAACC AGTACAAGGA GCTCCGGAAG

BamHI

6851 TTGGCCAGCT GCGCCTGCT GCGCGACGGC GAGCTGCTCA CCACCCAGGA
AACCGGTCGA CCGCGGACGA CGCGCTGCCG CTCGACGAGT GGTGGGTCCT

BamHI

6901 TCCGTCCCCC TTTTCCTTTG TCGATATCAT GTAATTAGTT ATGTCACGCT
AGGCAGGGGG AAAAGGAAAC AGCTATAGTA CATTAATCAA TACAGTGCAG

6951 TACATTACAG CCTTCCCCC ACATCCGCTC TAACCGAAAA GGAAGGAGTT
ATGTAAGTGC GGGAGGGGGG TGTAGGCGAG ATTGGCTTTT CCTTCCTCAA

7001 AGACAACTG AAGTCTAGGT CCCTATTTAT TTTTATATAG TTATGTTAGT
TCTGTTGGAC TTCAGATCCA GGGATAAATA AAAAAATATC AATACAATCA

7051 ATTAAGAACG TTATTTATAT TTCAAATTTT TCTTTTATTT CTGTACAGAC
TAATCTTTCG AATAAATATA AAGTTTAAAA AGAAAAAATA GACATGTCTG

7101 GCGTGTACGC ATGTAACATT ATACTGAAAA CCTTGCTTGA GAAGGTTTGT
CGCACATGCG TACATTGTAA TATGACTTTT GGAACGAACCT CTTCCAAAAC

HindIII

7151 GGACGCTCGA AGGCTTTAAT TTGCA
CCTGCGAGCT TCCGAAATTA AACGT


```

          X X      fs
          = =      =
1  MSITVTFPKS PSTTKRAPAF GIELEFSQOG ESDGATSKAA LAVPVFSVDN

      X X  X      R
      = =  =      =
51 QDFVLIRDLA KYWGYPSYQ LTVKLVKAN IEKSQILKTD KDLNKFEL
101 DLISEADTKI ELFYISLPLV YSRIENKKVF YVIREPEQPK VSKAPTQCKP
151 ASVVAASEDD ENLDDDEEDE VDZOMDEEDND MSGELSKGYK HHMKDHPKYI
201 NDDRVTIGQV FHQYGLDPST FLTHSLFNSI NSMSKLNYYK NFGVSGYRFL
251 PNSKLSYAEK ELVLNANNYN DMHINEKTES KPKKSFRKPI GKSKKHNLQI

          fs      T
          =      =
301 DFNSIDLSES VLPQGGFIPD FSIHMLCKVP NYVVTSNHQS LELSFNTKNL

          X
          =
351 NATSNSSYLF NENVKIKSKS IQKLVPNSDT DNYHHTKYFY TKTYRGPQSG
401 NYKDQALMNK INKIELESNK KPRHKRFVSN NWRYNKSLKG LVHEKFDKNF
451 VEYLLSEQRK YTEDYSNLEI LHNSLQFNVL LNTYRGVAQE TWNYYKFKL

          X      fs
          =      =
501 IDFEQLKALQ MEANELEERK LDAARHQOWA ESEKLRQERL RLVPEDEFNE

          X X  L  X  * X XX
          = =  =  =  =  =
551 FEQLQSEFGQ RKNDEEKLR RQLEASLSD SFEADSENDD ESELAQIQD

                                missing sequence
                                =====
601 FESSANALKT KFEAKYDLI NPAPPPQPIE TPQLDLNNKF SLPTVYPEIT

missing sequence
=====
651 PNLPLELRGV VPESXZELPP IKKAIHYVTT YPERPNXEYL TRNRDYPIAN

missing
=====
701 ANSGWXG

```

Fig 66

117c at

```
      L      Q
      =      =
1  MSCEDSEHNNH NKGHNHNNHH VAPIPI TAGQ SLNNKI DTSK VTALNMANSA

      ambiguities
      =
      Q
      =
51  DDLAKVFEDS TKKYPITK
```

Fig 67

15c1

```
                                fs  S
                                =  =
1  QQSYVZQSQP NYSQQTQDRG MFSGGGGGGHG HYQQQQGYN A YGPPPPQGGY

                                ambiguities
                                =====
                                X      W      W      WX
                                =      =      =      ==
51  YQQQPGGGGG YQQQQQQQP MIVQQQPRSG GNE SCLYGC L AALCVCTLD

      amb
      ==
101 MLF
```

Fig 68

222a8

```
      X      X      X      R      X
      =      =      =      =      =
1  MFDNFILKNL VDETESEIDS GETELSDDYY YYYSVELDCK EDDSDEITAQ

51  ILLENSELGT KTFNFEDPFE QINIEDNKVI EVNTPKTAKP TTTVGTSTTS

      X      XX
      =      ==
101 ALSTPESTIF EIPKFFYGSR RKQLSSPKNK NSTIKFDVFD WIFESGTTNE

151 KVNGLVLVSS GVLLGTCLLF IL
```

Fig 69

22209

```

      E
      =
1  MRRREIERRK QKXREQRQK EHEAKRDIRI QQLSEQDSRS NQTKXEEXVF
51  KXARSTNSGA DETGLMSDKE FDDAYSAPDY LFZENLWNKP NHPDTNHKTK
101 KYTENVVENL DDPFNDSAY NSSFHDETNI QNEIQIPEND EYVQMKATS
      K      D      VR  is  C
      *      =      ==  =  =
151 SVMNTTIPAQ RMESLSTSE NKRRXFETAD VGVKGLDSPX XAQRNIWKI
      P
      =
201 QVSDNPMTVY FFMXXRLET PEGKLLCRDQ

```

Fig 70

32801

```

      F
      =
ambiguities
==
1  MPRIXQVDVF TNYKYLGNPV AVIYDSNLT TQEMQKIARW TNLSETTFIL
51  TPKSSIADYS TRIFTSGONE LPPAGHPTLG TAFALLEJGK IKPNDNGQII
      ambiguities
      =====
101 QECGAGLVKI SVEKTPNWN NNNNNKSNEI PFLLSFELPY PKFHEIDDKV
      is
      =
151 IEELQHSWNG TMTIGKPVLI DAGPKWAVFQ LSGGKEVLDL NVDLAQIERL
      ambiguity
      ambiguities 2      G
      =====      =
201 SLENGWTGIG VEGHKNENGD SVELRNIAPA VGV

```

Fig 71

33gK part1

```

                                XY   XXXX           fs
                                ==   =====
1  NVTDSTPFAM GTLGSTFYAV TSVGRSFQIY DLATLHLLFV SQTQTFSRIT

      C   XF   X
      =   ==   =
51  SLAAHHHYVY ASYGDRIGIF RRGRLHELV CEGNSTVMQL LVFGELYLAT

                                X   X   XX           X   XXXX
                                =   =   ==           =   =====
101 TLEGDIFVFR KTEGKKFPE LYTTIRIINS LVEGEIVGLI HPPTYLNKVI

      XXX X
      === =
151 VATTQSVFVI NVRTCKLLYK SRELQFEGEK ISSIEAAPVL DVIAGVTSNG

                                fs
                                =
                                X   P
                                =   =
201 NVFLFNKKKG ESVGAKMLL LELNLFKVA SISFRTDGAP HLVAGLNNGD

                                A   I   R
                                =   =   =
251 LYFYDLKKK RVEVLNNAK ETHGCVANAK FLNGQPIVLS NGGDNHLKEP

      H   S           Y           G   R           C
      =   =           =           =   =           =
301 VFDPNLTTSN SSIVPPPRNL RSRGGHSAPP VAIEFPQEDK THFLLSASRD

      fs
      =
      Y           LC           ambiguities
      =           ==           =====
351 KTFWTFSLRK DAQAQEMSQR LQKSKDGKRQ AGQVVSMEK FPEIISISS

      amb
      ==
401 YAR
```

Fig 72

33gK part2

1 IITAKDETF APTWDSRNR VGRHLLNTID GGIVKSVCVS QCGNFGLVGS
51 SLGGIGSYNL QSOLLRKKYV LHKQAVTGLA IDGMNRKMVS COLDGIVGFY
101 DFGKSVYLGK LQLEAFITSM IYHKLSDLVA CALDDLSTIV IDVTQKQVIR
151 ILYGHTHRIS GMDPSPDGRW IVSVALDSTL RTWDLPTGGC IDGVILPIVA
201 TAVKFPSPIGD ILATTHVSGN GVSLWTNFAQ FKPVSTRHVE EDEFSTILLP
251 NASGDGGSTM LGGFLOEDSN EDGTIDEQYT SAAQIDASLI TLGSEPRSKF
301 NTLLKLDTIK QQSKPKKAPK KPENAPFFLQ LTG

Fig 73

33gK part3

1 KLRKLDITGN HAFSEFTKL LREAGISGQF ERFITYLLNL SPAVLDEIR
51 SLNSFVPLTE MTNFIQALNA GLKSNANYEI WETLYAMFPN IHGDVIHQFE
101 NETSLHEALE EYRQNDKKN NKQDSLKXYC ASIVSFIS

Fig 74

480g

	X	X		X	X	X	
	=	=		=	=	=	

1 MIQNNRADLD XVCNCWYDVY KLRKLDILNE SSSSISEQIH IRDRISRVYQ

	fsg X	Xfs	XG	FFXXE	V	S	DPTX
	==	==	==	=====	=	=	=====

51 PRILDLVRAI GIDKEEALKE KQLVSQIQES IDNLLVQEVV RSKRVLGGA

*KFGXXXIXQ*X X INV missing sequence
=====

101 KSTPETLPIN NKELLQHGVQ IHQNCDELD QLRVLIARQK QIGELINAEV

missing sequence
=====

151 EQQNEMLDRP NEEVDYTSSK IKQARRRAK IL

Fig 75

55q1

```

      NSX      MIs      XX G X N      R XFXW      X
      ====      =      == = = =      = =====
1  NLTLKYIHV KLSQLRNYL SHKSIETLEK MMSVNAVND BODLSSTSEI

      FG X R      GX X      R      X      XX X
      == = =      == =      =      =      == =
51 IRRFILEGVK SQTSTSKDIT SQQKLENFMT ILRETRPDEK VVEDYLIDVI

      F      * X      R
      =      =      =
101 APQIQLOSED YPDSVVLST PSIKGILSI MDSRNNANQI LLETRYGILL

151 KDANVPVLNK EDIVGCPDML SIGNPYGAKS NPPWLGTEI TQNGKNAGAN

      A
      =
201 NLLIEKLSVM TMCYESEILS SKLSFNAQDL DQESQENYND DNSKQAPLRL

      FV      IB
      ==      ==
251 GIDMPSVVIT STSSQYFTLY VIIVSLLFYS EPMSKVIHKK IEKMRPSIDF

      *      F      E YXX      X      * X
      =      =      = ===      = = =
301 EDLGALTSRL TMOQQHXLK KVLKXKXFP ERGNXTNEVL QFIYLQGESW

351 NGGE
```

Fig 76

60gK

FIG

===

1 ITDFSDFKIT KLPALAELE LKRCYICKEL LNAPVRTQCD HTYCSQCIIE
51 FLLRDNRCPL CKTEVFESGL KRDFLEELIV ISYASLRPHL LRLLEIEKVE
101 SKQEVDRREYS ANESALNGNR N/RINDVDET/ RVKDQLNADK LGEEKGQAQH
G fs X
= = =
151 WEQVNEQTTE VILLSDDEE NGSDSLVKCP ICFERMELDV LQGHIDDCL
fs
= =
Q Q X ambiguities
= = =
201 SGKSTKRTPT DILSPKAKRP KQITSFFKPT IDTKTPSPPT SKASTTPTAT
S Q N I K M
= = = = =
251 PTTTLKAW ASPSPVAQST VHKGKPLPKL DFSSLSQKI KAKLSDLKLP
* =
301 TTGSRNEMEA RYLNYYIYN ANLDSNHPV

Fig 77

8c gp

1 XQFSSAVVLS AVAGSALAAY SNSTVNDIQI TWVTITSCEE NKCHETEVTI
51 GVTTVTEVLT TYTTYCPLSI TEAPAPSTAT DVSTTVVITIT SCEEDKCHET
101 AVTTEVTVTV EGTIIYTYTC PLPSTEAPGP APSTABESKP AESSFVPTTA
151 AESSPAKTTA AESSPAQETT PKTVAAESSS AETTAPAVST AERGAANAV
201 PVAAGLLALA ALF

Fig 78

Fig 79

Fig 80

1 EPKVAKSXS TIGKIFRYTF YTAVISVIGS AGLIGYRIYE ESQPVQGVKQ

X
=

51 TPLFPNGEKK HTLVILGSGW GAISLLKNLD TTYNVVIVS PANYFLFTPL

fs X fs fs
= = = =

101 LPSVPTGTVE LRSIIEPVRS VTRRCPGQVI YLEASATNIN PKTNELTKQ

R N X X
= = = =

151 STTVVSGHSG KDTSSSKSTV AEVTGVEEIT TTLNYDYLWV GVGAQTILIF

X X X XX XX X
= = = = = = = =

201 GNPGRMRKF NPFFERKTSQ SHLQIR

Fig 81

3576L

1 MAXFIKAGKV AIVVGEYAG KKVVIVKEHD EGTKSHPPH AIVAGIERAP

51 LKVTXMDAX KVEKRTKVKP FVKLVNHNH MPTRYSLDVE SFKSAVTSEA

fs
=

101 LEEPSQREEA KVVKKAFEE KHQAGHWWF FQKLHF

Fig 82

7

1 MSIPSTQYGF FVXKASGLKL KADLPDNKPG AGQLLLKVDA VGLCKSDLEH

fs

=

X fs D X X X
= = = = =

51 LYEGLDCQDN VVMGHEIAGT VAEIGEEVSE FAVGERVACV GPMGCGLCKH

missing sequence

=====

ambiguities X XXXXIFGXHXX

=====

101 CLTGNDNVCT KSFLEWFGLG YNGGYEQILL VKRPRNLVKI PDNVTSEEA

missing sequence

=====

151 AITDAVLTFY HAKKSAGVGP ASNILIICAG GLGGNAIQVA KAFGAKVTVL

missing sequence

=====

201 DKHKDKARDQA KAFGADQVYS ELPDSVLPGS FSACFDFVSV QPTYDLCQKI

missing sequence

=====

251 CEPKGGTIVFV SLGATSLNIN LALDLREIT VKGSFWGTLM DLREAFELAA

missing sequence

=====

301 QGKVKPNVAH AFLSELPKYM EKLRAAGGYEG RVVFN

Fig 83

```

      M Y N      E      D
      = = =      =      =
1  MEKIDINTNS NKIQAYQKV VPGDENATFV VTSVDKNATM DVTETGDGSL

      Y      P      P  LSYD
      =      =      =  ====
51 EDFVEHFTDG QVQFGLARVI VPGSDVSKNI LLGWCPDSAP AKLRCAFANN

      S I      M N Y H X
      = =      = = = =
101 FADVSRVLSSG YHVQITARDQ DDLDVNEFLN RVGAAAGARY STQTSGLKKP

151 SPAAPKPTSK FVVAKSSAS KPSFVPKSTG KPVAPAKPKP KNITKDAGWG

201 DAEDVEERDF DKKPLDNVPS AYKPTKVNID ELRKQKSDTT SSTPKTFKSE

      P      HD R      T
      =      = =      =
251 PQEEKNDGDDG QSKPLSERMK AYDQPSSSDG RLTSLPKPKI GHSVADKYKA

      fs
      =
301 SASGNGAAPA FGAKPAFGTQ SVDSRKDKLV GGLSRDFGAE NGKTPAQIWA

351 EKRQKYNTVA DDEKETNSSE KVDEPSEHHA ADLAKKFEEL ANIAGDTPSL

      K      S fs X
      =      = = =
401 PTRNLPPAPP ARETAIPSNE KQKEEKZEEE QAPAPSLPTR NLPPFSORQP

      fs X      XS      missing sequence
      = =      ==      =====
451 EPEPEPEEEE EEEEEAPAP SLFARNLPPA PKAEAEESKK QSTTATAEYD

      missing sequence
      =====
501 YEKDEDNEIG FSEGDLIIDI EFVDDDMWQG KHAKTGVEVGL FPATYVSL

```

Fig 84

226a a12

```

1  VLGSQWDEG KGLVDLLCD DIDWCARCQG GNNAGHTIIV GKVKYDFHML

51  PSGLVNPCKQ NLVSGVVIH VPSFFAELEN LEAGLDCRD RLFVSSRAHL

      A      D TX
      =      = ==
101 VFDFHQRTDK LKEAELSTNK KSIGDTGKCI OPTYSTKASR SGIRVHHLVN

151 PDPEAWZEPK TRYLRVFSR QERYGEFEYD

```

Fig 85

P

1 EDKKQHFDAS GASAVDDKTA TAILRRKKKC NALVVDDATN DONSIVTMSS
 51 NIMELLQLFR GDTVLVWKK RKSTVLIVLA DDDMPDGVAR VNRVVRNLR
 101 VRLGDIVTVE PCPDIKYANR ISVLPIADTV EGINGSLEFL YLKPYPVEAY
 151 RZVRKGDLSF VRGGMRQVEF KVVEVDPEZI AIVAQDTIIR CEGEPINRED

R fs

201 FENSINEVGY DDIGGCKXQM AQIRSELVELP LRHPQLPKSI GIKPPKGILM
 251 YGPFCTCKT MARAVANETG AFFFLINGPE IMSXMAGESE SMLRKAPEEA

fs P fs N KK P
 = = = = =

301 EKNSPSIIFI DZIDSIAPKR DKTNGEVERF VVSQLLTMD GMRKARSXVY

G DQ L Pfs
 = == = ===

351 IAATRPNSI DPALRRPGRF DREVDIGVPD AGRLEILAI HTKNMKLADD
 401 VDLEAIASEF AGFVGADIAS LCSEAMQOI RERNDLIDLE EETICTEVLN
 451 SLGVIQDNER FALGNSNPSA LRETUVENVN VTWDDIGOLE NIKNELKZTV
 501 EYPVLHPDQY QXFG LAPTKG VLEFGPPGTG KOLLAKPVAT EVSANFISVK
 551 GFELLSMNYG ESESNERCIF EKARAAAPT VPLDELDSIA KARGGSHGDA
 601 GGASDRVVNQ LLETENDGMA KKNVFVIGAT NRPDQIDPAL LRFGRLDQII
 651 YVPLFDEPAR LSILQAQLRN TPLEPGLDLE BIAKITHGFS GADLSYIVQR

AfsQ I

= == =

701 SAKFAIKDSI EAQIKINKIK EEKEKVKTED VDMKVDSVEE EDPVPIYTRA
 751 HFEEAMKTAK RSVSDAELEZ YESYAOQLQA SRGQFSSFRF NENAGATONG

X

801 SAAGANSQAA FOX

Fig 86

89g3 part1

1 TLKQRLSEIL FAKAEVVKQF KKEHGKTVIG EVLLEQAYGG MRGKGLVWE
51 GSVLDPIEGI RFRGRTI2DI QKELPKAPGG EEPLPEALFW LLLTGEVPTD
101 AQTKALSEEF AAPSAIPKXV EELIDRSPSH LHPMAQFSIA VTALESSESQF
151 AQAYAKGANK SEIWXYTYED SIDLLAKLPT IAAKIYRNVE HDGKLFAAID
X T X T
= = = =
201 SKLDYGANLA ELLGFGDNKE FVELMRLYLT IESDHEGQNV SAHTTHLVGS
RI
= = =
251 ALSSPFLSLA AGLNGLAGPL HGRANQEVLE

Fig 87

89g3 part2

1 QREFALKHMP DYELFKLVSE IYEVAPGVLT KHGKTKNPF NVDSHSGVLL
S
=
51 QYYGLTEQSF YTVLFGVFRA FGVLPQLILD RGLGMPIERP KSFSTEKYIE
101 LVKNINK

Fig 88

408c5 part1

ambiguityX X ' GL
= = = =
1 SDYHVIMLAF RNNGIMEAEY RLYLLVITLI ISPVSXIMFG VGAAREWFWQ
51 VIYVGLGFIG FQNGSIGDTS MSYLMDAYPD IVIQGNVGVN IINNTLACIF
101 TFACSYWLDG SGTQNTYIAL SIIDFATIAL VFPFLYYGKT FRRKTKRLYV
151 SMVELTQGM

Fig 89

1 DQNNEDFIPG TENIYGLEVD SEDENVSHYD ASSRPKVTK GHILFPQPS
 51 NSCNDPLNWS KKRKLSNFFI VIFITAFTAA TSNDAGSIQD SLNEKYGISY
 101 DAMMTGACVL FLOIGWGTFF LTPASSLYGR KITYFICIFL GLLGAVWEAL
 151 VKSTSDSINS QLFVGISESC AEAQVQLSL ELVFAHNLGS VLTSYIVATS
 201 VGTYLGLPIA AFIVQNIGFR WVGWIAAIIIS GALLFVIVFC LDETYFDRAK
 251 FTKP

Fig 90

ESP1

1 DFQLQDILHH VESKWFGGFI SGIFTNDNDV ENESKNVTHK FKQDLMKILK
 51 DCLTVSDDKS NIEPFLQFNE FIYVCFYSME EYNYELVDDL IKFITINMNS

X

*

=

=

101 HGRIVNFGIN VELTKLHBLI KNLIDKVAKN KGRCDXQQK QQQQQQQQQQ

ambiguities

===

QQQQ fs

X

ambiguities

=====

=

=====

151 NSNNSQHTVL IFYANCNFP WESMEFLRK SISRMPSIHM LLDLVKSNTN

201 NKNKLMFVDK SNLYYLINPS GDLIRSENRF KKLFESENHLW RGEIGKLSSN

F

=

X

ambiguities

=

=====

251 EHEDYQDSIL CEIKSHLFV YIGHGGCDQY IKVSKLFKXC GRNQDLLNKL

301 PPSLLGCS VKLENCNIN NSSMLQPLN IYNWLNCKSS MILGNLNDVT

D

=

351 DKDIDIFTLS LQKRWGLIAU YNGSGHDYGM KKLDLTNCVV QSRSKCTLXY

401 LNGSAPVWYG LPX

Fig 91

A
=

1 STKGKVKHAT FESMVLKPD LKGIYAVGFE TPSAIQERAI MQTISGRDTI

M
=

51 AQAQSGGSKT ATFSIGMLEV IDTKSKSCQA LILSPTRELA IQIQNVVKEL

101 GGYMWINTHA SIGGHNVGED VKKLQGGQOI VSGTPERVID VIKRRNLQTR

151 NIKVLILDEA DELFTKGFKE QIYEIYKHLF PSVQVVVSA TLPREVLENT

fs T
= =

201 SKFTTDFVRI LVRDEISLL GIKQYVQCE REDWKFDTLG DLYDNLITIQ

251 AVIFCNFLK VWLADQMKK QNFTVAMHG DMKQDERDSI MNDFFRGNSR

301 VLISTDVNAR GIEVQVSLV INYDLPTDKE NYIHRIGRSG RFRKGTAIN

X
=

351 LITKDDVTEL VELEKYVSTK IKEMPMNINE IM

Fig 92

40c af

1 NVDTIITIT RFIQSGQIV APTATGELS LNALQFAK FIAMNIRRAE

51 LVNLIGVSGS ANSTGDVQXN LNVIGDEIFI NAMRSSNNVK VLVSEEQEDL

101 IVFRGGCTYA VOTTFIDGSS NIDAGVSVGT IFGVYKLQEG STGGISDVLR

missing sequence
=====

151 PGKENVAGY TMYGASAHLA LTTGHGVNLF TLCTQLGEFI LTHENLKLPD

missing sequence
=====

201 TKNIYSLNEG YSNKFPYVQ DYLDIKKEG YSLRYIGIMV ADVHETLLYG

missing sequence
=====

251 GIFAYTTLK RLYEFPMA LLMEQAGGSA VTIKGERILD ILPKGIHDKS

missing sequence
=====

301 SIVLOSKEV EXYIKHVPK

Fig 93

380c8

1 DNVSTSTAE ANXNRIKVKD EFPQEEQANT SLEDKPVSAV IGIIMCFI
51 AFGGFVFGPD TGTISGFIMX SDFLERFGGT KADGTLVFSN VRTGLMIGLF
X X X
= = =
101 NAGCAIGALF LSKWEDMYGR RVGINTAMIV YIVGIIVQIA SQHAWYQVMI
ambiguities
==
X
=

151 GRIITGLAVG XLSVLCPLFI SEVSPKHLRG TLVCCPQLMI TLGIFLOYCT
fs
=

201 TYGTRSYSDS RQWRIPLGLC FAWALCLVAG MVRMPESFRY LVGKDRIEDA
PR
==

251 KMSLAKTNKV SPEDPALYRE LQLIQAGVER ERLAGKASWG TLFNGKTKIF
IV missing sequence
== =====

301 BRVMLGVMLQ ALQQFNWOKN LFPSYLTSPX N

Fig 94

98c 62

missing sequence
=====

1 NAFVSGTITE FLVDVDATVE VGQELIKMEE GDAPAGGASA SEAPAKKEEA
missing sequence
=====

51 PEKAKEESAP AAPPKKEETK KEEPKKESKP APNKEESKKS TQSTTSAPT
missing sequence
=====

101 TNFSRNEERV KOTIRRLRIA ERLKESQNTA ASLTFNEVD MSNLMDFRKK
missing sequence
=====

151 YKDEFIKTG IKLGFMGAFS KASALALKET PAVNAAIENN DTLVFKDYAD
missing sequence X XX XX NX *
===== = == == == =

201 ISIAVATPKG LVTPVVRNAE SLSILGIEKE ISNLGKKARD GKLTLEDMTG
S X XX X C X X X F X F X IX
= = = = = = = = = = =

251 GTPTISNGGV FGSLYSTPII NMPQTAVLGL NGVKERPVTV NGQIVSRPM
301 YLALTYDHRV VDGFEAVIFL RTIKELIEDP RNMLL

Fig 95

249c af

1 ERMISGMVYN CLQKELETTT MSCPDVMDLY GSPRTREYKT TQEFIDAKYK
51 HLESPIGHVG KQAFMEYPIY FDYGFNTYLG DNFYSNYNLT ILDVSEVRIG
fs
=
101 NNVKCGPNVS ILIPTHEVDP TLRVDQLENA LPVTVGNGVN LOGSCTILGG
X missing sequence
=
151 VTUGDQSIVA AGAVVNDVP PNTVVAGVPA RVVKQLEPRD PNF

Fig 96

55a1

XX XXFY X P X X
== ===== = = = =
1 TSDTKTKQRE NPLKDISSQ GGNLRVPRS SSSSSSQKKK SSKKQAHNDE
T HL V L
= == = =
51 DDZENGCGEG FLDASSRRI LQLAKEQQDE LEQEDEIQNK P3FAQSFKNO
D L L FI S
= = = == =
101 QIDSEEEEEE DEYSDTEEBE EVEEIVDEE DAEVDPKDAE LFNKYFQNG
D FX R X P X F
= == = = =
151 EANNDEEDNE FQPTINLADK ILAKIQEKES QQQQQQSSP DNSNEDAVLL
E I
=
201 PPKVILAYKR IGQILSTYTH GXLPKLFKIL PSLKNWQDVL YVTNPNSTF
251 HATYEATKLF VSNLSSNEAT VPTETILLPR FRDSIENSDD HSLNYHIYRA
301 LKKSLYKPGA FPKGFLLFLV DGYOSVREAT IAASVLTKVS VPVLHSCRYC
351 GVLNNKPKES FVEVLRI

Fig 97

1 MYVYKRDGRK EPVRFDKITA RVQRLCYGLN PNHVEPVAIT QKVISGVYQG
 31 VTTIELDNLA AELAAATMTI HPDYAVLAAR IAVSNLHKQT TKQYSKVSKD
 101 LYEVINPKTG LHSPMISKET YDILMEHEDZ LNSAIVYDRD FNNYVFCFKT
 151 LERSYLLRIN GKVAERPQEL IMRVAVGICG NDIPRVIETY NLMSQRFFTH
 201 GSPCLFNAGT FRPQMSSCFL LAMKDDSIETG IYDTLKSCAL ISKSAGGIGL
 251 HIHNIRSTGA YIAGTINGTSN GIIPMVRVFN NTARYVDQGG NKRPGAFALY
 301 LEPWHSIDFD FIDIRKNHGK BEIRARDLFP ALWIPDLFMK RVEQNGDWTL
 351 FSPNEAPGLA DVGDEFEEL YTRYEKENRG RQTIKAQKLW YAILGAQTET
 401 GTFFMLYKDS CNKSNQKNL GIIKSSNLCC EIVEYSAPDE VAVCNLASIA
 451 LPSFVENDER STWYNFKLH QVTKVVTNL NRVIDRNHYP VPEAERSNMR
 501 HRPALGVQGG LADAFMEIRL PFDSQEARL NIQIFETIYH AAVEASIELA
 551 KEEGAYETYP GSPASQGLLQ FDLWNRKFTB LWDWDTLKQD LAKHGMRNSL
 601 LVAPMPTAST SQILGNNECF EPYTSNIYSR RVLAGEFQIV NPYLLDLVD
 651 LGVWNDAMKS SIALWNGSIQ ALPNIPDEIK ALYKTVWEIS QKHIIDMAAD
 701 RAAFIDQSQS LNEHDKDPTM GKLTSMHFYG WKKGLKTGMY YLRTQAASAA
 751 IQFTIDCKIA ETAGHTVANL DKLNIKXYVN KGRVESENTS DAPYKSFSTE
 801 PLSLESSVAD LKXDEGEKEF AEDKTIEELE NDIYSAKVIA CAIENPESCT
 851 MCSC

Fig 98

485cl

	fs	X	fs		V
	=	=	=		=

1 APKMYQSEDV PAPKQTRKTA RPQKLRLASLA PGTVLILLAG RFRGKRVVYL
 51 KNLEDNTLLV SGFFVWVGV LRPVNAVYVI ATSTKVNVSG VDVSKFNVEY
 101 FAREXSSKX XSEAEFFNES QPKKEIKAER VADQKSVDAA LLSEIKKTPL
 151 LKQVLAASF LKNGRPHLL KF

Fig 99

328c3

1 MENDKGQIVE LYVPRKCSAT NRIIKAKDEA SVQISIARVD EDGRAIAGEN
51 ITYALSGYVR GRGEADDSL N LAQQDGLLK NVWSTSR

Fig 100

67q1 part1

X
=
1 MSPKGFKKGV LAAPQTNRQE FNMGEITQDA VYLDARERFK BIEMETKKLS
51 EESKKYFNAV NGMLDEQIDF ARAVAELYKP ISGALSDPSA TVFEDNPQGI

P
=
101 EASELYQAVV KDLKDTLCPD LELIEKRIVE LAQELLKIIQ AIRKMS

Fig 101

67q1 part2

A
=
1 HKRNFSKYEL KKERTVXDEE KMFSAQTEVE IAQGEYDYNN DLLKNELFVL
Y Y I D
=
51 FQMQSDFIKP LFVSYFMQL NIFYTLYTRM BELKIPYFDL STDIVEAYTA
S fs fs M
=
101 KKGNIIEQTD AIGITHFKVG HAKSKLEATK RRHAAMNSPP PTGASSIAST
151 GTGELPAYS PGGYNQPYGD SKYQPPSSPA TYQSPVVAAT AQSPATYQSP
201 VATGQPPSYL PQTPASAPFP QVGSGLPTCT ALYDYTAQAQ GDLTFPAGAV
251 IELIORTEDA NGWMTGKYNG QTCGVFPNGYV QL

Fig 102

1 MTTSKZETFLF TSSEVIGZGHE DKICDQVEDA ILDACLAVIDF LSKVACETAA
 51 KTGXIMVFESE ITTKAQLDYQ KIIFDTIKHI GYDRSEKGFQ YKTCNVIVAI
 101 EQQSPDIAQG LHYEKALBEL GAGDQGIMFG YATDETDEKL PATILLAKKL
 151 NAALASARSS GSEFWLRPDT KTQWTIEYEK DGGAVIPKRV DTIVISTQHA
 201 EEITPENLFX EPIEHIIKQV IPEHLLDDKT IYHIQPSGRF VIGGPQGEAG
 251 LTGRKIIIVDT YGGWGAHGCG AFSCKDFSKV DRSAAYAARN VAKSLVTAGL
 301 AKRALVQFSY AIGVAEPTSI VIDTYGTSLK STEALVEIHK NNFDLRPGVI
 351 VKELELARPV YFKTASYGHE TNQENSWEQP KKLKF

Fig 103

2329 CG

1 PKNSADAKYT MEILRGIRLY GKALKLKRID AXSQSSTNRP NNQTIGTFVQ
 51 SDLINPNYID VGAKLFNNL NPLVDESFLM DTFSKFGTLI RNPIIRDSSE
 101 GHSLGYGFLT YDDFESSDLC IQKQNTILM NNKIALSYAF KDLSVQGRKS
 missing sequence
 =====
 151 RHGDQVERKL GKXNP

Fig 104

83c3

```

ambiguities      L  XP  W  L      S  M
=====
1  CPFLTEISLH LKVMQHONIA TIHQVLNIED FAIILMDHF EQGDLEFNII

      S X      X
      = =      =
51  DRQIFTNNSH RKVPFTDFET QLLMKNAMLQ LIEAIEYCHE NNIVHCDLKP
101 ENIMVRYNPN YVRPTINNNW NNGEDDLQYA NSIIDYNELH LVLIDFGLAM
151 DSATICCNHC RGESFYMAPE RTTNYNTHRL INQLIDMNY ESIEINGTTV

                                F Y
                                = =
201 TKSNCXYLPT LAGDIWSLGV LFINITCSRN PWFIASFOHN QNNEVIKXNM

      L  V  L      L      D  D
      =  =  =      =      =  =
251 MNMNAKLSK IIPISSQFNR ILDRIFKLNK NGRIGL

```

Fig 105

233c cd2

```

1  NGTKQQITSI IDMLNKADLP KEVEVVICPP ALVLGLAVEQ NKQPTVAIGA

                                X
                                =
51  QNVFDKSCGA FTGETCASQI LDVGASWTLT GHSERRTIK ESDEFIAEKT

      X      X  X      GCWFQDXX      XGX
      =      =  =      =====      ---
101 KPALDTGVKV ILCIGETLEE RKGGVTLQVC ARQLDAVSKI VSDNSNIWVA

                                missing sequence
                                =====
      LKNWVWFYFX XKMLKKPTRX*T  X  X XXX
      =====
151 YEPVMEIGTG LAATPEDARE THKGIRAHLA KSIGAEQAER TRILYGGSVN

      missing sequence
      =====
201 GKNVXDFKDK AWDGFLVGG ASLKFEFVDI IKSR

```

Fig 106

This Page Blank (uspto)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)